Methodological Considerations in School Climate Research

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

Yuane Jia, M.A.

August 2019

**Dissertation Abstract**

Methodological issues in school climate research have received somewhat less attention in the literature. These issues include the impacts of failing to screen data for potentially invalid responses in self-reports of school characteristics, the measurement issue of school climate as a level-two organizational construct, the common informant-based effect across different rater types (e.g., students and teachers), and the boundary and mechanism of how and why school climate influences student and school outcomes. These issues are addressed in a series of studies that comprise this three-paper dissertation. All three studies drew upon data from the statewide assessment of school climate surveys and department of education records in Virginia public high schools.

The first paper focused on the effects of validity screening in self-report data by investigating associations between students' bullying victimization experiences and a series of student adjustment outcomes. Two methods of identifying invalid responders were used: survey completion time and built-in validity check items. Results revealed that inclusion of the invalid responders in the total sample inflated the prevalence of all reported risk behaviors, and deflated student reports of GPA, school engagement, and depression. The contrasts between the valid and invalid respondent groups were statistically significant ($p < .001$), with meaningful effect sizes for all investigated outcomes, after controlling for student and school demographics. The invalid group was significantly comprised of more males, non-Whites, and younger students. The associations between bullying and student outcomes varied as a function of group membership (valid vs. invalid). Results supported the study hypotheses that validity screening significantly affected the summary statistics/prevalence rates of self-reported outcomes and also impacted the

associations among variables. This study provided additional evidence for the need to use validity screening in self-report surveys of adolescents.

The second paper investigated when and how different components of school climate, reported by students and/or teachers, were associated with school-level dropout rates. We attempted to examine how different elements of an authoritative school climate (i.e., disciplinary structure, academic expectations, and student support) were related to high school dropout rates, whether the components of school climate interact with each other to predict school dropout rates, and whether these relations were influenced by differences between student and teacher informants. The study found that student reports of teachers' academic expectations, teacher reports of their support for students, and the interaction between the two, were significantly associated with school dropout rates. Results further indicated that high academic expectations were most strongly related to school completion when teacher-student relationships were supportive, and that both high academic expectations and supportive student-teacher relationships were associated with lower dropout rates even in schools with a high proportion of low-income students. The work provided an example of using appropriate statistical procedures to reveal complicated relationships in social science studies.

In the third paper, we illustrated the use of a doubly latent multilevel structural equation modeling framework (MSEM) in applications of mediation analysis, by examining the mediating role of student engagement in associations between school climate ratings obtained by teachers and a variety of student outcomes, both self-reported outcomes and external records. The illustrations demonstrated how to apply the relatively new procedure--doubly latent MSEM approach into school climate research with real data in applications of multilevel mediation analysis, while controlling for both measurement error and sampling error and providing bias-

corrected estimates of the indirect effects. Substantive results supported the hypothesis that student engagement mediates the relationship between school climate reported by teachers and student outcomes of Prevalence of Teasing and Bullying, GPA, and suspension rates. It provided a functional model to explain why and how school climate affects student outcomes. This can inform school improvement efforts and provide evidence for the development of effective interventions. Together, this three-paper dissertation provided a window into several important methodological considerations in studying school climate and call for more attention and research on appropriate use of methodology in social science in general.

Curry Programs in Research, Statistics and Evaluation
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation ("Methodological Considerations in School Climate Research") 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.

Name of Chair (Dr. Timothy Konold)

Committee Member (Dr. Dewey Cornell)

Committee Member (Dr. Patrick Meyer)

Committee Member (Dr. Michael Hull)

Date

**Acknowledgements**

This dissertation is based on research conducted throughout my five-year doctoral program journey. This work would not have been possible without the financial support of the Youth Violence Project and department of Education Leadership, Foundations and Policy at the University of Virginia Curry School of Education. I am especially indebted to Dr. Timothy Konold, the chairman of my committee, and Dr. Dewey Cornell, director of Youth Violence Project, who have been supportive of my career goals and who worked actively to provide me with guidance, patience, time, finance, and space to pursue those goals. As my teachers and mentors, they have taught me more than I could ever give them credit for here. They have shown me what a good scientist, teacher, leader and person should be.

I am grateful to all of those with whom I have had the pleasure to work during this and other related projects. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance, thank you also to  Dr. Patrick Meyer, Dr. Michael Hull, and Dr. Ji Hoon Ryoo. Other key members I wish to thank are Francis Huang, Jennifer Maeng, Kathan Shukla, Pooja Datta, Kelly Edwards, Marisa Malone, Anna Grace Burnette, Brittany Crowley, Shelby Stohlman, and Caroline Crichlow-Ball. I wish to acknowledge each and every faculty/staff member and classmates of the Curry School of Education Research, Statistics and Evaluation Program.

Finally, I would like to give special gratitude to my loving and supportive husband, Silin, and my three young children, Jason, Josiah, Joshua, who provided unending inspiration. Without you I would not have the courage to complete the journey with these accomplishments. I would also like to thank my parents, parents-in-law, my sisters, and my friends for their support.

## TABLE OF CONTENTS

## Project Overview

A large and growing body of research points to the importance of a positive school climate on a variety of student outcomes. As a contextual factor, school climate is related to a variety of student outcomes (Jones & Shindler, 2016). For example, school climate has been found to be strongly associated with student academic, behavioral, and social-emotional development (Cohen, McCabe, Michelli, & Pickeral, 2009; Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013; Wang & Degol, 2016; Wilson, 2004). Because school climate is a malleable factor that can promote positive student development and bring about meaningful change in student well-being, it has become a central target and catalyst for school reform efforts (Wang & Degol, 2016).

The focus of the current set of papers is on methodological issues in school climate research that have received somewhat less attention in the literature. These issues include the impacts of failing to screen data for potentially invalid responses in self-reports of school characteristics, the measurement issue of school climate as a level-two organizational construct, the common informant-based effect across different rater types (e.g., students and teachers), and the boundary and mechanism of how and why school climate influences student and school outcomes. These issues are addressed in a series of studies that comprise this three-paper dissertation. The first paper focused on the effects of validity screening in self-report data by investigating associations between students' bullying victimization experiences and a series of student adjustment outcomes. The second paper investigated when and how different components of school climate, reported by students and/or teachers, were associated with school-level dropout rates. In the third paper, we illustrated the use of a doubly latent multilevel structural equation modeling framework in applications of mediation analysis, by examining the mediating role of student engagement in associations between school climate ratings obtained by teachers and a variety of student outcomes, both self-reported outcomes and external records.

**Paper One**

There are many benefits to using anonymous self-reports in school settings, such as cost-efficiency, less interviewer bias (Fan et al., 2006), and protection of participant privacy and confidentiality. Anonymity is presumed to facilitate more honest responses on sensitive personal experiences (e.g., health-risk behaviors; Brener, Billy, & Grady, 2003). However, anonymity may also hold the potential of reducing accountability and decreasing respondent motivation to answer questions carefully and truthfully (Douglas & McGarty, 2001; Lee, 2006; Lelkes, Krosnick, Marx, Judd, & Park, 2012; Meade & Craig, 2012). Concerns over truthfulness and accuracy of self-report data might be most warranted in surveys that address sensitive issues (e.g., substance use, victimization experience, sexual orientation) and/or target young adolescent respondents (Cornell, Klein, Konold, & Huang, 2012; Fan et al., 2006; Pokorny, Jason, Schoeny, Curie, & Townsend, 2001). For instance, adolescents have been found to both intentionally under- and over-report difficult-to-recall and sensitive risk behaviors due to social desirability beliefs (Brener et al., 2003). Failure to detect and screen these potential invalid respondents from self-report data prior to analysis is likely to contaminate substantive conclusions.

Although most school climate studies do not explicitly screen data for potential invalid respondents, research has found that even a small proportion of invalid responders can compromise study findings (Cornell et al., 2012; Fan et al., 2002, 2006; Robinson & Espelage, 2011). Examples include studies that illustrate that inclusion of invalid responses consistently results in inflated prevalence rates of risk behaviors, and that valid responders have more positive perceptions of their schools (Cornell et al., 2012; Cornell, Lovegrove, & Baly, 2014).

A large body of research has documented the associations between bullying experience and student adjustment outcomes across academic and social-emotional domains (e.g., Beran & Li, 2008; Juvonen & Graham, 2014; Lacey & Cornell, 2013; Schneider, O'Donnell, Stueve, & Coulter, 2012; Smalley, Warren, & Barefoot, 2016; Swearer, Espelage, Vaillancourt, & Hymel, 2010; Van Geel, Vedder, & Tanilon, 2014). Most of the evidence linking bullying to student adjustment is based on anonymous self-reports that are not screened for potential invalid responses. The purpose of paper one was to investigate the hypothesis that validity screening would impact the accuracy of summary statistics and prevalence rates of student self-reported adjustment outcomes, as well as the associations between student's bullying experiences and a series of student academic and social emotional outcomes. Although the impact of validity screening on prevalence rates has been studied by others (e.g., Cornell, et al., 2012) the impact on associations between constructs has received little attention.

The sample was obtained from the 2014 Virginia Secondary School Climate Survey (VSSCS; Cornell, Huang, et al., 2014). The VSSCS is a statewide survey of school climate and safety conditions in Virginia public secondary schools, which was administered to 323 out of 324 eligible public high schools in the spring of 2014. The school participation rate was 99.7%. It 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.

Bullying victimization was measured with a five-item scale that included global, physical, verbal, social, and cyber 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). Student academics outcomes included self-reports 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 (e.g., Bauman, Toomey, & Walker, 2013; Kowalski & Limber, 2013) as well as the association between bullying experiences and student adjustment (e.g., Reed, Nugent, & Cooper, 2015).

Two methods of identifying invalid responders were used. First, the survey included two validity screening items: "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, the survey completion time was examined to identify surveys that were completed in an improbably brief time.

Results revealed 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. The contrasts between the valid and invalid respondent groups were statistically significant ($p < .001$), with meaningful effect sizes for all investigated outcomes, after controlling for student and school demographics. The invalid group was significantly comprised of more males, non-Whites, and younger students. Group membership, identified by validity screening items and survey completion time, was found to play a moderating role in the relationship between bullying and all student outcomes (all *p*s $< .05$) with the exception of student self-reported depression ($p > .05$). That is, the association between bullying and student outcomes varied as a function of group membership, and the magnitudes of the differences between the valid and invalid groups were dependent on the level of reported bullying victimization. The larger group differences were observed for higher levels of bullying victimization experiences. Results supported the study hypotheses that validity screening significantly affected the summary statistics/prevalence rates of self-reported outcomes and also impacted the associations among variables. This study provided additional evidence for the need to use validity screening in self-report surveys of adolescents.

Paper one, "Jia, Y., Konold, T. R., Cornell, D., & Huang, F. (2018). The impact of validity screening on associations between self-reports of bullying victimization and student outcomes. *Educational and Psychological Measurement,78*(1), 80-102 . " was published.

**Paper Two**

Perceptions of school climate tend to vary across different informants (e.g., students and teachers) as a consequence of their different roles within the schools. This raises the methodological question of whether one informant type is a better source for gauging the quality of schools climate than another. A goal of paper two was to examine which measures of school climate, and by which informants, were most strongly associated with high-school dropout rates after controlling for school demographics. In other words, we sought to examine how different elements of an authoritative school climate (i.e., disciplinary structure, academic expectations, and student support) were related to high school dropout rates, whether the components of school climate interact with each other to predict school dropout rates, and whether these relations were influenced by differences between student and teacher informants.

Dropout rates in high schools are a national concern (Freeman & Simonsen, 2015). Several studies have investigated individual risk factors associated with dropping out of school (Freeman & Simonsen, 2015; Legters & Balfanz, 2010) as well as school-level factors that are predictive of dropout rates. Some examples include proportions of low income students, school size, and school urbanicity (Balfanz & Legters, 2004; Freeman & Simonsen, 2015; Rumberger, 2011). However, student and school demographics are often static factors that are difficult to change. By contrast, school climate is a malleable factor that might ameliorate demographic risk and increase student motivation and engagement in school, thereby lowering dropout rates (Freeman & Simonsen, 2015; Legters & Balfanz, 2010; Rumberger, 2011).

School climate is a multidimensional construct that has been measured in a variety of ways by different researchers. Authoritative school climate provides a useful theoretical framework for understanding school climate that is derived from theories of authoritative parenting. Here, school structure and student support are analogous to the concepts of demandingness and responsiveness used in the parenting literature (Gregory & Cornell, 2009; Gregory et al., 2010). Structure has both behavioral and academic elements: disciplinary structure and academic expectations. Disciplinary structure is defined as the degree to which school rules are perceived as strict but fairly enforced and academic expectations refers to the degree to which teachers demand high academic performance in their students. Support refers to the degree to which students feel supported and respected by their teachers, and are willing to seek help from them.

In the parenting literature, authoritative parenting (i.e., high in both demandingness and responsiveness) has been found to be the most effective parenting style that leads to positive development outcomes and less maladjustment outcomes in children (Larzelere, Morris, &

Harrist, 2013). Likewise, recent authoritative school climate research has found that the interaction between structure and support was related to the prevalence of teasing and bullying in schools (Cornell et al., 2015) and suspension rates in high school (Gregory, Cornell, & Fan, 2011).

Authoritative school climate has important qualities associated with positive student outcomes and reducing negative outcomes. However, the multidimensional nature of authoritative school climate leaves uncertainty as to which elements are most important in the role of reducing dropout rates, and whether some of these associations (e.g., support and dropout rates) vary as a function of other authoritative school climate elements (e.g., academic expectations; Lee, 2012; Gregory et al., 2011) and informant type (i.e., student vs. teacher). Because examinations of interactions may be more important than studying the association of a single factor with dropouts, and may more closely approximate the complexity of schools (Luyten, Visscher, & Witziers, 2005), our statistical models included several interactive product terms.

The sample drew from Virginia Secondary School Climate Survey (Cornell, Huang, et al., 2014), a statewide assessment of school climate and safety conditions in Virginia public secondary schools. A total of 323 of 324 (99.7%) high schools participated in the survey. Eight alternative schools comprising students at risk for dropping out, and identified as being statistical outliers with respect to dropout rates, were removed from the sample. In addition, dropout rates were unavailable for one school, leaving a final analytic sample of 314 schools with student reports were provided and 301 schools with teacher reports.

School climate was measured by two key components: structure and support. Student reports of school climate included two scales to measure school structure: Disciplinary Structure and Academic Expectations, and one scale to measure responsiveness or supportiveness of teacher-student relationships, Student Support. The teacher survey only included Disciplinary Structure and Student Support without Academic Expectations because teachers tended to endorse the highest level across all the items. The psychometric qualities of these scales were established in the context of a multilevel structural analysis of a more comprehensive authoritative school climate measurement model (Konold & Cornell, 2015; Huang & Cornell, 2016).

School dropout rates were obtained from the Virginia Department of Education (VDOE) for the 2009 and 2010 cohorts that graduated in 2013 and 2014, respectively. Cohort dropout rates were calculated as the percentage of student dropouts within a longitudinal 4-year cohort group (VDOE, 2014). Average dropout rates for the 2009 and 2010 cohorts within each school served as the outcome in the current study, which we believe it more reliable than any single year of the dropout rates for each school (Cornell, Gregory, Huang, & Fan, 2013; Lee, Cornell, Gregory, & Fan, 2011). Other school demographic variables were obtained from the VDOE website.

Two-step hierarchical regression analyses indicated that school-level demographic variables (i.e., school size, percentage of free and reduced-priced meals (FRPM)) were found to explain 38% of the variance in dropout rates; student reports of teachers' academic expectations, teacher reports of their support for students, and the interaction between the two, were significantly associated with school dropout rates, together explaining an additional 7.5% of the variance in school dropout rates beyond the school-level covariates. Further investigation of the interaction effect between support and academic expectations revealed that dropout rates were lowest in schools characterized by high student support and high academic expectations, whereas the level of academic expectations did not show a material difference in adjusted dropout rates when student support was low. Schools with high levels of FRPM were associated with highest dropout rates; however, schools characterized by higher levels of student support and academic expectations have decreased dropout rates.

This study examined how student and teacher perceptions of school climate could be combined to provide a more comprehensive assessment of school climate and how the elements of school climate measures interact with each other to exert their effects on school-level outcomes (i.e., dropout rates). The work provided additional support for the authoritative school climate theory in that results revealed that students were more responsive to teacher demands when they perceive them as concerned and respectful (Gregory & Cornell, 2009). Results further indicated that high academic expectations were most strongly related to school completion when teacher-student relationships were supportive, and that both high academic expectations and supportive student-teacher relationships were associated with lower dropout rates even in schools with a high proportion of low-income students.

Paper two, "Jia, Y., Konold, T. R., & Cornell, D. (2016). Authoritative school climate and high school dropout rates. *School Psychology Quarterly,* 31, 289-303." was published.

**Paper Three**

In contrast to paper two that examined the moderating role of variables in understanding linkages between measures of school climate and student outcomes, paper three attempted to adopt a relatively new approach to examine the potential mediating role of student engagement in understanding relationships between measures of school climate and student outcomes in multilevel contexts. Moderating effects are typically represented by product terms in models for purposes of evaluating potential interaction effects. Consequently, they are useful for understanding questions concerning “for whom, under what condition, or when does the relationship exist.” (MacKinnon & Luecken, 2008). Mediation analyses, however, focus on the processes by which variables are related to one another. For example, student perceptions of support in the school have been shown to have a direct relationship with student achievement (Berkowitz, Moore, Astor, & Benbenishty, 2017; Cornell, Shukla, & Konold, 2016). However, it’s quite possible that student support influences student engagement, which in turn produces greater achievement gains due to students being more specifically involved in the types of activities that foster learning. Here, the indirect effect of support on achievement through engagement provides a clearer picture of the mechanism through which support exerts its influence on achievement. The third paper investigated the mediating role of student engagement in the associations between teacher reports of school climate and a variety of student outcomes through recently developed methods of doubly latent multilevel structural equation modeling (MSEM; Marsh et al., 2009; Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Lüdtke et al., 2008). This approach overcomes limitations in traditional multilevel modeling (MLM) in applications of mediation analysis.

Several advantages of using this doubly latent MSEM approach in school climate research have been described. First, it provides better control over sampling and measurement error than is afforded by traditional observed variable multilevel models and reduces bias in substantive effect estimates. School climate is typically evaluated by individual informants (e.g., student or teacher ratings) residing within schools. However, more than individual experiences (Cohen et al., 2009), school climate is a school-level construct that reflects collective experiences, shared beliefs, values, and attitudes among the members within each school. From multilevel perspectives (Marsh et al., 2012; Wang et al., 2014), the individual (Level 1: L1) responses (e.g., student or teacher ratings) include sources of variance that are shared with other school informants and are representative of true L2 trait variance, and residual variance sources that are unique and unrelated to the L2 trait being measured (Marsh et al., 2012). Failure to disaggregate residual variance components in L2 measures may introduce bias in estimates of group-level effects (Lüdtke et al., 2008; Shin & Raudenbush, 2010).

The majority of school climate studies use cluster means by manifestly aggregating individual ratings to represent an L2 construct. This approach may introduce bias by ignoring sampling error (i.e., only sampling a subset of individuals in each school) (Lüdtke et al., 2008; Shin & Raudenbush, 2010). This is more problematic when an unequal number of individuals are sampled in each school (Wang & Degol, 2016). In a similar vein, the use of scale means as L2 construct measures could introduce bias in the indirect/direct effect estimates by ignoring measurement error due to sampling of items (Muthén, Muthén, & Asparouhov, 2016). The doubly latent MSEM approach results in a truer measure of school climate and less bias in substantive effect estimates by simultaneously correcting for measurement error through the use of multiple indicators at both levels, and sampling error through latently aggregating individual responses into L2 constructs (Lüdtke et al., 2011; Marsh et al., 2012; Marsh et al., 2009).

Second, MSEM can easily decompose the within and between components of the L1 variables as latent, which has been shown to be dramatically less biased and to provide better confidence interval coverage in applications of mediation analysis (Preacher, Zhang, & Zyphur, 2011). Traditional MLM conflates the between and within effects by reporting a single average slope. If the within effect and between effect are different (e.g., a contextual effect exists; Raudenbush & Bryk, 2002; Marsh et al., 2009), the single slope estimate will misrepresent the data and result in a biased inference for either level. In fact, for any mediation model involving a variable assessed at L2, the indirect effect can only strictly exist at the cluster level (Preacher, Zyphur, & Zhang, 2010). Accordingly, proper disaggregation of the total L1 variable variance into between and within components is important (Arens, Morin, & Watermann, 2015; Morin, Marsh, Nagengast, & Scalas, 2014). Otherwise, the indirect effect will be biased if not properly estimating the path between variables assessed at L1 because they should be separated into two parts; one is strictly at the between level and the other strictly at the within level.

Third, the MSEM framework is adaptable to accommodate any mediation designs (i.e., any outcomes and mediators assessed at either level), add multiple mediators, or allow random slopes serving as mediators or independent variables (Preacher et al., 2010). Finally, MSEM can provide fit indices to evaluate model fit (Preacher et al., 2010).

The usefulness of doubly latent MSEM in the context of school climate research is illustrated by investigating whether student engagement mediates the associations between school climate and a variety of student outcomes, after controlling other school characteristics (e.g., %FRPM, %White, school size). We obtained measures of school climate (i.e., fairness, justness, and support) through teacher ratings, and measures of engagement and student outcomes of teasing and bullying (PTB) and self-report GPA in schools through student ratings. Different informant types were used to help alleviate potential concerns with shared method variance (Wang & Degol, 2016) and common informant-based effects (Konold & Cornell, 2015). Since school climate is conceptualized as a school-level construct, primary interest was in the L2 indirect effects of engagement on relationships between school climate and student outcomes.

The initial sample consisted of 62,679 student responses from 320 schools and 12,249 teacher responses from 302 schools. Data screening included identifying survey response times that were too fast, failing built-in validity check items (see Cornell, Huang et al., 2016 for additional details), removing schools with less than 3 teacher responses, and schools with reports from only students or teachers. The resulting analytic sample consisted of 59,581 students and 11,336 teachers from 296 schools. The student sample (51.3% female) was distributed across 9th (27.5%), 10th (25.9%), 11th (24.6%), and 12th (22.0%) grades. The race/ethnicity breakdown of students was 54.5% White, 17.6% Black, 11.8% Hispanic, 5.9% Asian, 1.0% American Indian or Alaska Native, 0.5% Native Hawaiian or Pacific Islander, with an additional 8.7% reporting more than one race. Teacher respondents were predominantly female (67.0%) and were 83.1% White, 8.3% Black, 3.4% Hispanic, 1.3% Asian, 0.2% American Indian or Alaska Native, 0.1% Native Hawaiian or Pacific Islander, and 3.4% were classified as other/two or more races.

Information from VDOE records indicated that the 296 schools had an average enrollment of 1,397 students (range 59 to 4,190), and the average percentage of students eligible for FRPM in the participating schools was 38.9%, with a range of 1.5% to 100.0%. The average percentage of White students across all the participating schools was 59.6%, with a range of 0.4% to 100.0%.

The primary measures used in this study were obtained from the Virginia Secondary School Climate Survey (Cornell, Huang, et al., 2016), which has revealed favorable psychometrics at both individual and school levels (e.g., Huang, et al., 2015; Konold et al., 2014). The teacher reports of authoritative school climate include three factors: justness, fairness and student support at the school level; these havedemonstrated good psychometric properties in a multi-level CFA framework (Huang et al., 2015; Huang & Cornell, 2016). Student self-reports of engagement has two components, affective and cognitive engagement. A previous study (Konold et al., 2014) with a statewide middle school sample demonstrated preferable psychometric properties in multilevel contexts. Student outcomes of Prevalence of Teasing and Bullying (PTB) have demostrated good validity in a series of substantive studies (e.g., Huang, Cornell, & Konold, 2015; Konold et al., 2014; Konold & Cornell, 2015).

Doubly latent MSEM was used to specify the multilevel mediation models, see the graphic representation of the model in figures 1-3. A two-step approach was used to examine the amount of school-level outcome variance that could be explained by the school climate and engagement factors beyond the school covariates. In step 1, paths a1-a3 (the mediator student engagement was regressed on the climate factors), c1-c3 (the outcome was regressed on the climate factors), and b (the outcome was regressed on the mediator engagement) were constrained to zero in order to assess the amount of variance in outcomes that could be explained by the covariates. Thereafter in step 2, all paths were freely estimated. The indirect effects (product of a and b) of the school climate factors on PTB through Engagement were measured to assess the potential processes underlying these linkages, and a Monte Carlo (MC) based parametric bootstrap approach was adopted to evaluate the significance of the mediation effects. A weighted least squares means and variance adjusted (WLSMV) was selected as the estimator because most of the indicators can be considered categorical (Beauducel & Herzberg, 2006; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Analyses were implemented through Mplus 7.3 in the context of a MSEM framework. All school-level factors are doubly latent in the sense of simultaneously correcting for measurement error and sampling error due to the sampling of items and persons when aggregating L1 ratings into L2 constructs (Marsh et al., 2012; Morin et al., 2014).

The illustrations demonstrated how to apply the doubly latent MSEM approach into school climate research with real data in applications of multilevel mediation analysis, while controlling for both measurement error and sampling error (Marsh et al., 2009, 2012), and provide bias-corrected estimates of the indirect effects (Preacher et al., 2010). They showed the flexibility of this approach by accommodating different outcomes assessed at L1 or L2 with single or multiple indicators. The Mplus syntax used in generating the results of our models is provided as an Appendix to facilitate interested readers use of this procedure. Substantive results, based on carefully controlled for potential bias, supported the hypothesis that student engagement mediates the relationship between school climate reported by teachers and student outcomes of PTB, GPA, and suspension rates. It provided a functional model to explain why and how school climate affects student outcomes. This can inform school improvement efforts and provide evidence for the development of effective interventions.

**Implications**

School climate is a broad and complex construct and it exerts impact on student and school outcomes in complicated ways. This three manuscript dissertation aimed to provide a window into several important methodological considerations in studying school climate, including the need for validity screening, the use of different informants, and how to better aggregate L1 ratings to form L2 constructs.The third and culminating paper examined moderation effects to show how school climate components interact with each other to shape student outcomes, tested mediation mechanisms of how school climate impacts a variety of student outcomes in multilevel contexts, and demonstrated how to address potential bias in indirect effect estimates with nested data, adjusting for shared method variance in how different informants may influence outcomes. This work holds promise for improving the reliability and validity of school climate measures, minimizing estimation bias in the associations among variables, facilitating a deeper understanding of how and when school climate exerts its influences on student/school outcomes, and ultimately guiding school improvement and intervention efforts.

## References

Arens, A. K., Morin, A. J. S., & Watermann, R. (2015). Relations between classroom disciplinary problems and student motivation: Achievement as a potential mediator? *Learning and Instruction*, *39*, 184–193. https://doi.org/10.1016/j.learninstruc.2015.07.001

Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis: Which high schools produce the nation's dropouts? Where are they located? Who attends them?* Retrieved from http://files.eric.ed .gov/fulltext/ED484525.pdf

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*(2), 142–163. https://doi.org/10.1037/1082-989X.11.2.142

Bauman, S., Toomey, R. B., & Walker, J. L. (2013). Associations among bullying, cyberbullying, and suicide in high school students. *Journal of Adolescence*, *36*(2), 341–350. https://doi.org/10.1016/j.adolescence.2012.12.001

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*, 186–203. https://doi.org/10.1207/s15328007sem1302_2

Benbenishty, R., Astor, R. A., Roziner, I., & Wrabel, S. L. (2016). Testing the causal links between school climate, school violence, and school academic performance: a cross-lagged panel autoregressive model. *Educational Researcher*, *45*(3), 197–206. https://doi.org/10.3102/0013189X16644603

Berkowitz, R., Moore, H., Astor, R. A., Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research, 87*, 425–469. doi:10.3102/0034654316669821

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

Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, *111*(1), 180–213. Retrieved from https://schoolclimate.org/climate/documents/policy/School-Climate-Paper-TC-Record.pdf

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., Huang, F., Datta, P., Malone, M., Jia, Y., Burnette, A.G., Shukla, K., Konold, T., & Meyer, P. (2016). *Technical Report of the Virginia Secondary School Climate Survey: 2016 Results for 9th–12th Grade Students and School Staff.* Charlottesville, VA: Curry School of Education, University of Virginia.

Cornell, D. G., Lovegrove, P. J., & Baly, M. W. (2014). Invalid survey response patterns among middle school students. *Psychological Assessment*, *26*(1), 277–287. https://doi.org/10.1037/a0034808

Cornell, D., Gregory, A., Huang, F., & Fan, X. (2013). Perceived prevalence of teasing and bullying predicts high school dropout rates. *Journal of Educational Psychology*, *105*(1), 138–149. https://doi.org/10.1037/a0030416

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment*, *24*(1), 21–35. https://doi.org/10.1037/a0024824

Cornell, D., Shukla, K., & Konold, T. (2015). Peer victimization and authoritative school climate: A multilevel approach. *Journal of Educational Psychology*, *107*(4), 1186–1201. https://doi.org/10.1037/edu0000038

Cornell, D., Shukla, K., & Konold, T. R. (2016). Authoritative school climate and student academic engagement, grades, and aspirations in middle and high schools. *AERA Open*, *2*(2), 2332858416633184. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/2332858416633184

Douglas, K. M., & McGarty, C. (2001). Identifiability and self-presentation: Computer-mediated communication and intergroup interaction. *British journal of social psychology, 40*(3), 399-416.

Freeman, J., & Simonsen, B. (2015). Examining the impact of policy and practice interventions on high school dropout and school completion rates: a systematic review of the literature. *Review of Educational Research*, *85*(2), 205–248. https://doi.org/10.3102/0034654314554431

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

Gregory, A., & Cornell, D. (2009). "Tolerating" adolescent needs: moving beyond zero tolerance policies in high school. *Theory Into Practice*, *48*(2), 106–113. https://doi.org/10.1080/00405840902776327

Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for black and white high school students. *American Educational Research Journal*, *48*(4), 904–934. https://doi.org/10.3102/0002831211398531

Gregory, A., Cornell, D., & Fan, X. (2012). Teacher safety and authoritative school climate in high schools. *American Journal of Education*, *118*(4), 401–425. https://doi.org/10.1086/666362

Gregory, A., Cornell, D., Fan, X., Sheras, P., Shih, T.-H., & Huang, F. (2010). Authoritative school discipline: High school practices associated with lower bullying and victimization. *Journal of Educational Psychology*, *102*(2), 483–496. https://doi.org/10.1037/a0018562

Huang, F. L., & Cornell, D. G. (2016). Multilevel factor structure, concurrent validity, and test–retest reliability of the high school teacher version of the authoritative school climate survey. *Journal of Psychoeducational Assessment*, *34*(6), 536–549. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/0734282915621439

Huang, F. L., Cornell, D. G., Konold, T., Meyer, J. P., Lacey, A., Nekvasil, E. K., … Shukla, K. D. (2015). Multilevel factor structure and concurrent validity of the teacher version of the authoritative school climate survey. *Journal of School Health*, *85*(12), 843–851. https://doi.org/10.1111/josh.12340

Huang, F. L., Cornell, D. G., & Konold, T. R. (2015). Aggressive attitudes in middle schools: A factor structure and criterion-related validity study. *Assessment*, *22*(4), 497–512. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1073191114551016

Jia, Y., Konold, T. R., & Cornell, D. (2016). Authoritative school climate and high school dropout rates. *School Psychology Quarterly*, *31*, 289–303. https://doi.org/10.1037/spq0000139

Jones, A., & Shindler, J. (2016). Exploring the school climate-student achievement connection: making sense of why the first precedes the second. *Educational Leadership and Administration*, *27*, 35. Retrieved from http://search.proquest.com/openview/6f4931eed47a6226718977eb6ebe30c2/1?pq-origsite=gscholar&cbl=2032315

Klein, J., Cornell, D., & Konold, T. (2012). Relationships between bullying, school climate, and student risk behaviors. *School Psychology Quarterly*, *27*(3), 154–169. https://doi.org/10.1037/a0029350

Konold, T., Cornell, D., Huang, F., Meyer, P., Lacey, A., Nekvasil, E., … Shukla, K. (2014a). Multilevel multi-informant structure of the Authoritative School Climate Survey. *School Psychology Quarterly*, *29*(3), 238–255. https://doi.org/10.1037/spq0000062

Konold, T., Cornell, D., Huang, F., Meyer, P., Lacey, A., Nekvasil, E., … Shukla, K. (2014b). Multilevel multi-informant structure of the Authoritative School Climate Survey. *School Psychology Quarterly*, *29*(3), 238–255. https://doi.org/10.1037/spq0000062

Konold, T. R., & Cornell, D. (2015). Measurement and structural relations of an authoritative school climate model: A multi-level latent variable investigation. *Journal of School Psychology*, *53*(6), 447–461. https://doi.org/10.1016/j.jsp.2015.09.001

Kowalski, R. M., & Limber, S. P. (2013). Psychological, Physical, and Academic Correlates of Cyberbullying and Traditional Bullying. *Journal of Adolescent Health*, *53*(1, Supplement), S13–S20. https://doi.org/10.1016/j.jadohealth.2012.09.018

Lacey, A., & Cornell, D. G. (2016). School administrator assessments of bullying and state-mandated testing. *Journal of School Violence*, *15*(2), 189–212. https://doi.org/10.1080/15388220.2014.971362

Larzelere, R. E., Morris, A. S. E., & Harrist, A. W. (Eds.). (2013). *Authoritative parenting: Synthesizing nurturance and discipline for optimal child development*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/ 13948-000

Lee, H. (2006). Privacy, publicity, and accountability of self-presentation in an on-line discussion group. *Sociological Inquiry*, *76*, 1–22. doi: 10.1111/j.1475-682X.2006.00142.x

Lee, J. S. (2012). The effects of the teacher-student relationship and academic press on student engagement and academic performance. *International Journal of Educational Research*, *53*, 330–340. http://dx.doi.org/10.1016/j.ijer.2012.04.006

Lee, T., Cornell, D., Gregory, A., & Fan, X. (2011). High suspension schools and dropout rates for Black and White students. *Education & Treatment of Children*, *34*, 167–192. http://dx.doi.org/10 .1353/etc.2011.0014

Legters, N., & Balfanz, R. (2010). Do we have what it takes to put all students on the graduation path? *New Directions for Youth Development, 2010,* 11– 24. http://dx.doi.org/10.1002/yd.359

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. https://doi.org/10.1016/j.jesp.2012.07.002

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods, 16*(4), 444–467. https://doi.org/10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203–229. https://doi.org/10.1037/a0012869

Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, *16*, 249–279. http://dx.doi.org/10 .1080/09243450500114884

MacKinnon, D. P., & Luecken, L. J. (2008). How and for whom? Mediation and moderation in health psychology. *Health Psychology*, *27*(2, Suppl), S99–S100. https://doi.org/10.1037/0278-6133.27.2(Suppl.).S99

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764–802. https://doi.org/10.1080/00273170903333665

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Mehta, S. B., Cornell, D., Fan, X., & Gregory, A. (2013). Bullying climate and school engagement in ninth-grade students. *Journal of School Health*, *83*(1), 45–52. https://doi.org/10.1111/j.1746-1561.2012.00746.x

Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: an illustration. *The Journal of Experimental Education*, *82*(2), 143–167. https://doi.org/10.1080/00220973.2013.769412

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: the Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology*, *31*(5), 629–656. https://doi.org/10.1080/01443410.2011.586416

Pokorny, S. B., Jason, L. A., Schoeny, M., Curie, C. J., & Townsend, S. M. (2001). Eliminating Invalid Self-Report Survey Data. *Psychological Reports*, *89*(1), 166–168. https://doi.org/10.2466/pr0.2001.89.1.166

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: the advantages of multilevel sem. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(2), 161–182. https://doi.org/10.1080/10705511.2011.557329

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

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*(Supplement C), 128–137. https://doi.org/10.1016/j.childyouth.2015.05.016

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354-373. http://dx.doi.org/10.1037/a0029315

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*(7), 315–330. https://doi.org/10.3102/0013189X11422112

Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge, MA: Harvard University Press. http://dx.doi.org/10.4159/harvard .9780674063167

Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics*, *35*(1), 26–53. Retrieved from http://www.jstor.org/stable/25654102

Shukla, K., & Konold, T. (2017). A Two-Step Latent Profile Method for Identifying Invalid Respondents in Self-Reported Survey Data. *The Journal of Experimental Education*, 1–16. https://doi.org/10.1080/00220973.2017.1315713

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. http://dx.doi.org/10.3102/0013189X09357622

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357–385. Retrieved from http://journals.sagepub.com/doi/abs/10.3102/0034654313483907

Tofighi, D., West, S. G., & MacKinnon, D. P. (2013). Multilevel mediation analysis: The effects of omitted variables in the 1-1-1 model. *British Journal of Mathematical & Statistical Psychology*, *66*(2), 290–307. https://doi.org/10.1111/j.2044-8317.2012.02051.x

Van Geel, M., Vedder, P., & Tanilon, J. (2014). 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 Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing Group-Level Constructs From Individual-Level Survey Data. *Organizational Research Methods*, *12*(2), 368–392. https://doi.org/10.1177/1094428107309322

Virginia Department of Education. (2014). *Graduation, completion, dropout & postsecondary data*. Retrieved from http://doe.virginia.gov/statistics_reports/graduation_completion/index.shtml

Wang, M.-T., & Degol, J. L. (2016). School climate: a review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wang, W., Vaillancourt, T., Brittain, H. L., McDougall, P., Krygsman, A., Smith, D., … Hymel, S. (2014). School climate, peer victimization, and academic achievement: Results from a multi-informant study. *School Psychology Quarterly*, *29*, 360–377. https://doi.org/10.1037/spq0000084

Wilson, D. (2004). The interface of school climate and school connectedness and relationships with aggression and victimization. *Journal of School Health, 74*, 293–299

## Abstracts

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

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.

### Manuscript Two: Authoritative School Climate and High School Dropout

This study tested the association between school-wide measures of an authoritative school climate and high school dropout rates in a statewide sample of 315 high schools. Regression models at the school level of analysis used teacher and student measures of disciplinary structure, student support, and academic expectations to predict overall high school dropout rates. Analyses controlled for school demographics of school enrollment size, percentage of low-income students, percentage of minority students, and urbanicity. Consistent with authoritative school climate theory, moderation analyses found that when students perceive their teachers as supportive, high academic expectations are associated with lower dropout rates.

### Manuscript Three: Moving to the Next Level: Doubly Latent Multilevel Mediation Models with School Climate Illustrations

Traditional observed variable multilevel models for evaluating indirect effects are limited by their inability to quantify measurement and sampling error that can result from the sampling of observed variables and persons within level two units, respectively. They are further restricted by being unable to fully separate within- and between-level effects without bias. Doubly latent multilevel mediation models reduce these biases by decomposing the observed within-level indicators into within- and between-level latent components. This decomposition simultaneously controls for both measurement error at each level and sampling error in the aggregation of individual values that serve as indicators of higher-level constructs. We illustrate the usefulness of this approach in investigating the mediating role of engagement in associations between reports of school climate and a variety of student outcomes (i.e., self-report GPA, Prevalence of Teasing and Bullying, suspension rates). Analyeses are based on a statewide high school sample of 59,581 students and 11,336 teachers from 296 schools. Results reveal that student engagement serves as a mediator between school climate factors and student outcomes.

**Manuscript One**

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

Yuane Jia, Timothy R. Konold, and Dewey Cornell

Curry School of Education, University of Virginia

Francis Huang

College of Education, University of Missouri

Published, *Educational and Psychological Measurement*

Author Note

Yuane Jia and Timothy R. Konold, Department of Educational Leadership, Foundations, and Policy, Curry School of Education, University of Virginia; Dewey Cornell, Curry School of Education, University of Virginia; Francis Huang, College of Education, University of Missouri.

We thank members of the project research team including Patrick Meyer, Kathan Shukla, Pooja Datta, Anna Grace Burnette, Marisa Malone, and Anna Heilbrun Catizone. This project was supported by Grant #2012-JF-FX-0062 awarded by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice; and Grant #NIJ 2014-CK-BX-0004 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect those of the Department of Justice, the funding agency.

Correspondence concerning this article should be addressed to 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

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*: self-report surveys, validity screening, bullying victimization, risk behavior, academic adjustment, high school students

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

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, statewide 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 underand 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 Strongly 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 3 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 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 *p*s < .001) and in expected directions. Higher reports of bully victimization were associated with lower GPAs ($B$ = 20.03), lower affective ($B$ = 20.15) and cognitive ($B$ = 20.05) 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 *p*s < .001), and group membership was also found to play a moderating role in the relationship between bullying and all student outcomes (all *p*s < .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 for the continuous outcomes. For example, a one-unit increase on the bully victimization scale yields a decrease in affective engagement of 0.15 ($B_B$ = 20.15) 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., ($B_B$ = 20.15) + ($B_{BxV}$ =0.14) = 20.01). 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 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 and 4. Here again, group membership was found to have a statistically significant effect on all reported outcomes (all *p*s < .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 (*p*s < .05), and the use of only response time played a moderating role in 7 of the 10 student outcomes (*p*s < .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, 2017); 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.

## 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. (2016). ‘‘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*, *28*, 458-470. http://dx.doi.org/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 selfinjury 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. (2017). Effects of mischievous responding on universal mental health screening: I love rum raisin ice cream, really I do! *School Psychology Quarterly, 32, 320-335*. 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-psych010213-115030

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. (2017). A two-step latent profile method for identifying invalid respondents in self-reported survey data. *The Journal of Experimental Education*, 1-16. Advance online publication.

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

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 |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *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 | | Chi-square |
|---|---|---|---|---|---|
| | *N* | % | *N* | % | |
| 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% | |

*** $p < .001$.

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*** |
| $\Delta R^2$ | .01*** | .03*** | .01*** | .10*** |
| Step $3_{\text{SI and RT}}$ | | | | |
| Invalid (I) | -0.24*** | -0.75*** | -1.07*** | -0.20*** |
| B x I | 0.03** | 0.14*** | -0.12*** | -0.01 |
| $R^2$ | .12*** | .16*** | .09*** | .19*** |
| $\Delta R^2$ | .01*** | .01*** | .03*** | .01*** |
| Step $3_{\text{SI}}$ | | | | |
| Invalid (I) | -0.29*** | -0.81*** | -1.11*** | -0.22*** |
| B x I | 0.00 | 0.24*** | -0.02 | -0.05*** |
| $R^2$ | .12*** | .16*** | .09*** | .19*** |
| $\Delta R^2$ | .01*** | .01*** | .03*** | .01*** |
| Step $3_{\text{RT}}$ | | | | |
| Invalid (I) | 0.13*** | -0.41*** | -0.93*** | -0.25*** |
| B x I | 0.02 | -0.02 | -0.37*** | 0.13*** |
| $R^2$ | .11*** | .15*** | .07*** | .18*** |
| $\Delta R^2$ | .00 | .00 | .01*** | .00 |

Note. Invalid = responders identified as invalid group (valid group as reference), B x 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 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 |
| $R^2$ ($\Delta R^2$) | .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 x 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 |
| $R^2$ ($\Delta R^2$) | .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 x 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 |
| $R^2$ ($\Delta R^2$) | .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 x 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 |
| $R^2$ ($\Delta R^2$) | .15***(.00) | | .13***(.00) | | .05***(.00) | | .07***(.00) | | .13***(.01***) | | .16***(.00) | |

Note. Invalid = Responders identified as invalid group (valid group as reference), B x I = Bullying by invalid interaction, SI = invalid responders identified through screening items, RT = invalid responders identified through response time, OR = Odds Ratio.
**p < .05. **p < .01. ***p < .001.*

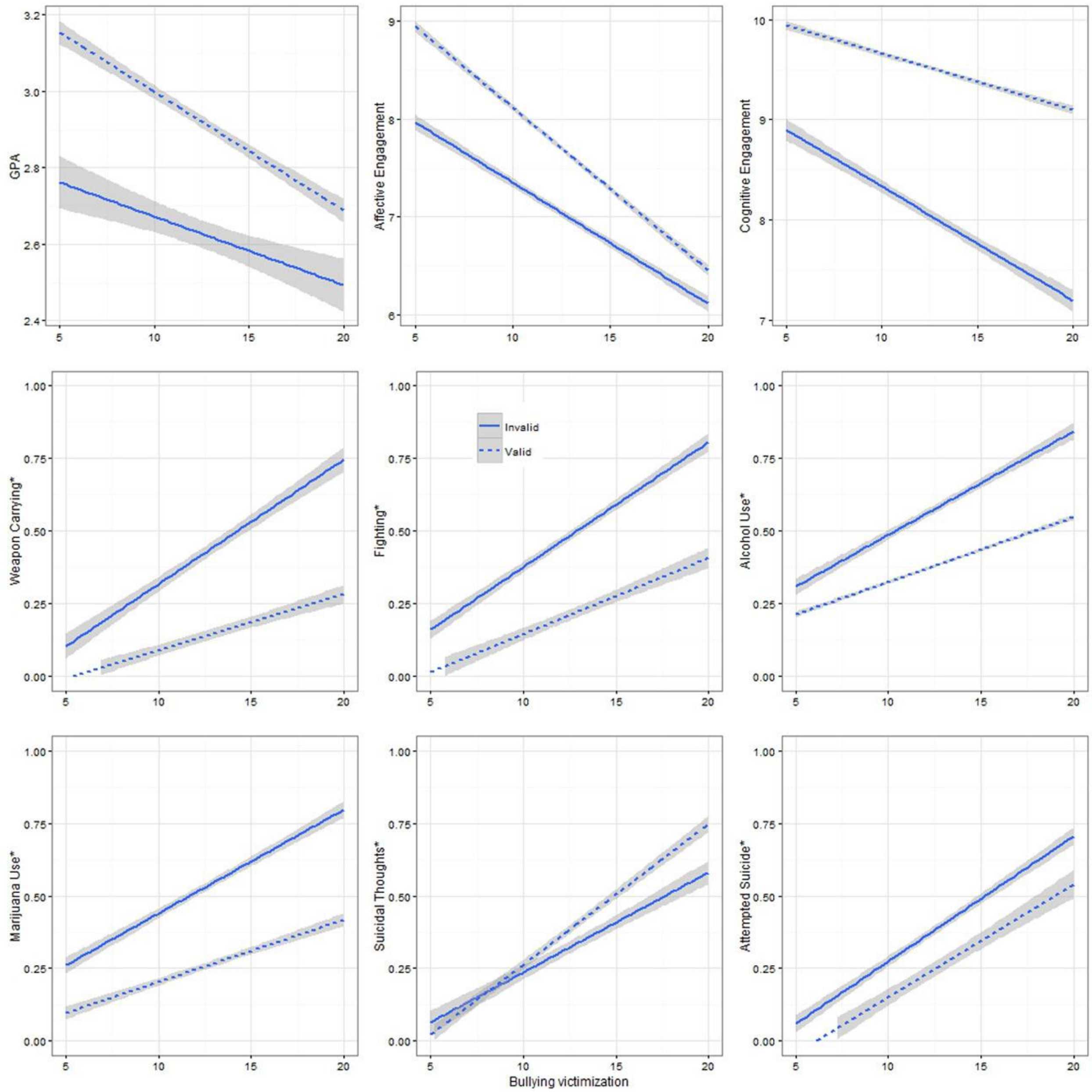


*Figure 1.* Associations between bully victimization and the predicted student outcomes by valid and invalid group membership.
*Note.* *Indicates dichotomous outcomes.

**Manuscript Two**

# Authoritative School Climate and High School Dropout Rates

Yuane Jia, Timothy R. Konold, and Dewey Cornell

Curry School of Education, University of Virginia

Published, *School Psychology Quarterly*

Author Note

Yuane Jia and Timothy R. Konold, Department of Educational Leadership, Foundations, and Policy, Curry School of Education, University of Virginia; Dewey Cornell, Curry School of Education, University of Virginia.

This project was supported by Grant 2012-JF-FX-0062 awarded by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice. We thank Donna Michaelis and Jessica Smith of the Virginia Department of Criminal Justice Services for their support of the Virginia Secondary School Climate Study.

Correspondence concerning this article should be addressed to 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

Abstract

This study tested the association between school-wide measures of an authoritative school climate and high school dropout rates in a statewide sample of 315 high schools. Regression models at the school level of analysis used teacher and student measures of disciplinary structure, student support, and academic expectations to predict overall high school dropout rates. Analyses controlled for school demographics of school enrollment size, percentage of low-income students, percentage of minority students, and urbanicity. Consistent with authoritative school climate theory, moderation analyses found that when students perceive their teachers as supportive, high academic expectations are associated with lower dropout rates.

*Keywords*: academic expectations, high school dropout rates, moderation, school climate

## Authoritative School Climate and High School Dropout Rates

High school dropout rates are a serious national problem (Freeman & Simonsen, 2015). One in five public high school students fails to graduate on time (Stetser & Stillwell, 2014), and approximately seven percent do not obtain a high school diploma or its equivalent by age 24 (National Center for Education Statistics, 2014). Students dropping out of high school are more likely to be unemployed and receive public welfare, and they are at increased risk for mental health problems, gang membership, and criminal behavior (Belfield & Levin, 2007; Rumberger, 2011; Swanson, 2009). Although there is a large body of research on individual student risk factors for dropping out, it is important to consider school-level factors, too, because approximately 12% of schools produce nearly half of the nation's students who drop out of school (Balfanz & Legters, 2004). Consequently, a large body of research has examined how high schools can lower their dropout rates (Freeman & Simonsen, 2015; Legters & Balfanz, 2010).

The demographics of a high school are consistently associated with its dropout rates. Schools with higher proportions of low income students, typically measured by the percentage eligible for free or reduced-price meals, have consistently higher dropout rates (Balfanz & Legters, 2004; Christle, Jolivette, & Nelson, 2007; National Center for Education Statistics, 2015; Rumberger, 2011; Suh & Suh, 2007). Students from low income families experience many more stressful events and difficult circumstances that make it difficult to attend school, engage in learning, and be successful. Research also indicates that the percentage of nonwhite students in schools is significantly associated with higher dropout rates (Christle et al., 2007; Stetser & Stillwell, 2014), and that larger schools tend to have higher dropout rates (Lee & Burkam, 2003; Leithwood & Jantzi, 2009; Werblow & Duesbery, 2009). Finally, many studies report that schools in urban areas have exceptionally high dropout rates, perhaps because of higher rates of poverty and other risk factors (Balfanz & Legters, 2004; Swanson, 2009). The causal relations among these factors are complex, but the overall effect is that some student populations face social and economic disadvantages that make high school completion more difficult, and consequently, schools with higher proportions of students with these demographic characteristics have higher dropout rates (Balfanz & Legters, 2004; Freeman & Simonsen, 2015; Rumberger, 2011). Although student demographics are static factors that schools cannot change, much research has focused on characteristics of the school climate that might ameliorate demographic risk and increase student motivation and engagement in school (Freeman & Simonsen, 2015; Legters & Balfanz, 2010; Rumberger, 2011).

**School Climate and Dropout**

School climate is a multidimensional construct that has been broadly defined as the "quality and character of school life" and "based on patterns of people's experiences of school life and reflects norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures" (Cohen, McCabe, Michelli, & Pickeral, 2009, p. 182). A broad definition of school climate makes it difficult to identify specific features of the school climate that are critical to positive outcomes (Cornell & Mayer, 2010; Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013; Wang & Degol, 2015). Currently, there are more than 30 instruments designed to assess various dimensions of school climate (American Institutes for Research, 2013). Many of these instruments consist of a diverse series of scales that are not linked by a larger conceptual model.

Authoritative school climate theory is a conceptual model derived from Baumrind's research on authoritative parenting that stimulated a large body of child development research

(Baumrind, 1968; Larzelere, Morris, & Harrist, 2013). Parenting research has found that authoritative parents exhibit two general qualities, often labeled demandingness and support. Demandingness is present when parents demand strict discipline and have high achievement expectations for their children and support is indicated by emotional responsiveness and warmth. In contrast, demanding but not supportive parents are described as authoritarian, supportive but not demanding parents are permissive, and those who are neither supportive nor demanding are neglectful or indifferent. Across a variety of studies, children of authoritative parents tend to have higher levels of pro-social behavior and better school adjustment, as well as lower levels of aggression, anxiety, depression, and substance abuse (Larzelere et al., 2013).

Authoritative school climate theory applies the framework of authoritative parenting to schools. Pellerin (2005) used administrator and student survey data to classify high schools as authoritative based on measures of demandingness and responsiveness. Demandingness was represented by separate student and administrator perceptions of two qualities: school rules are strictly enforced and academic expectations are high. Responsiveness was based on separate student and administrator perceptions that teachers are warm and nurturant toward students. Pellerin found that authoritative high schools had less truancy and fewer dropouts than other schools, controlling for school demographics of student socioeconomic status, school size, and urban location.

Lee (2012) measured authoritative school climate through a student survey using academic press (high academic expectations for students) as an indicator of demandingness and teacher– student relationship (teachers get along well with students and treat them fairly) as an indicator of responsiveness. This study found that high academic press and supportive teacher–student relationships were associated with higher behavioral and emotional student engagement in high school and that a supportive teacher–student relationship, but not academic press, was predictive of reading performance.

The Authoritative School Climate Surveys were developed to measure characteristics of an authoritative school climate using student informants (Konold et al., 2014) and teacher informants (Huang et al., 2015). This model uses the terms “structure” and “support” to refer to the two key dimensions of school climate that are analogous to the concepts of demandingness and responsiveness used in the parenting literature (Gregory & Cornell, 2009; Gregory et al., 2010). Structure has both behavioral and academic elements, that is, disciplinary structure and academic expectations. Disciplinary structure is defined as the degree to which school rules are perceived as strict but fairly enforced and academic expectations (sometimes called academic press) refers to the degree to which teachers demand high academic performance in their students. Support refers to the degree to which students feel supported and respected by their teachers, and are willing to seek help from them. Structure and support may have both direct and interactive effects on student outcomes. For example, it has been hypothesized that students are more responsive to high demands from their teachers when they perceive them as concerned and respectful (Gregory & Cornell, 2009). In other words, the effect of structure on student outcomes may be moderated by student support. A study by Gregory, Cornell, and Fan (2011) determined that the interaction of high academic expectations and student reports of support was associated with lower suspension rates for high school students. In particular, students were better behaved and less likely to be suspended in schools where students perceived both high support and high academic expectations from their teachers.

Authoritative school climate theory does not claim that structure and support encompass all aspects of school climate, but there is considerable evidence that they are important qualities associated with positive student outcomes. Johnson's (2009) review of 25 studies concluded that "schools with less violence tend to have students who are aware of school rules and believe they are fair" and "have positive relationships with their teachers" (p. 451). Two studies found that disciplinary structure and student reports of support were associated with less peer aggression such as teasing and bullying (Gregory et al., 2010; Cornell, Shukla, & Konold, 2015). Notably, Cornell, Shukla, and Konold (2015) found that a significant interaction between disciplinary structure and student support for both the prevalence of teasing and bullying at school and for student reports of bullying victimization. One interpretation of these findings was that students were more willing to comply with school expectations for respectful peer interactions when they perceived their teachers as supportive. Other studies found that disciplinary structure and student support were associated with lower levels of student aggression toward teachers, but did not find interaction effects (Berg & Cornell, 2016; Gregory, Cornell, & Fan, 2012).

Academic expectations is used in this study as an indicator of school structure or demandingness to represent an authoritative school climate, but other studies have examined it on its own (Goddard, Sweetland, & Hoy, 2000; Lee & Smith, 1999). These studies have concluded that a school climate that places a high value on learning will stimulate greater student achievement (Goddard et al., 2000). In contrast, low teacher expectations negatively affect student outcomes and school effectiveness (Brault, Janosz, & Archambault, 2014). In one study, 69% of students dropping out of high school had indicated that adults did not expect them to perform well, and that these low expectations played a role in their decision to drop out of school (Bridgeland, Dilulio, & Morison, 2006). Suh and Suh (2007) found that students' educational expectations had a significant impact on their decision to continue high school regardless of their at-risk status (e.g., low SES, academic difficulties, or behavior problems). Bridgeland, Dilulio, and Balfanz (2009) concluded that teacher expectations of high academic standards for all students can play an important role in reducing dropout rates.

In the aggregate, there is a substantial body of work supporting the idea that authoritative school climates are associated with lower school dropout rates. However, the multidimensional nature of authoritative school climate leaves uncertainty as to which elements are most important in the role of reducing dropout rates, and whether some of these associations (e.g., support and dropout rates) are moderated by other authoritative school climate variables (e.g., academic expectations; see Lee, 2012). Examinations of moderating influences through interactions may be more important than studying the association of a single factor with dropouts and may more closely approximate the complexity of schools (Luyten, Visscher, & Witziers, 2005).

A related and important consideration is that perceptions of school climate tend to vary as a function of different informants. Although students and teachers may share objectively similar experiences, their perceptions of those experiences often vary significantly (Mitchell, Bradshaw, & Leaf, 2010). Research (Kearney & Peters, 2013; Mitchell et al., 2010) has revealed low associations between teacher and student perceptions of classroom climate and overall school climate, but few studies have evaluated their differential associations with important school-level outcomes (Konold & Cornell, 2015).

**The Present Study**

The purpose of the current study was to examine how different elements of an authoritative school climate, including disciplinary structure, academic expectations, and student support, are related to high school dropout rates, and to evaluate whether these relations were influenced by differences between student and teacher informants. Furthermore, this study considered whether the associations between authoritative school climate and dropout rates exist independently of school demographics. This work was focused on three guiding research questions: (a) What are the associations between student perceptions of disciplinary structure, academic expectations, and student support and school dropout rates after controlling for school demographics? (b) What are the associations between teacher perceptions of disciplinary structure and student support and school dropout rates after controlling school demographics? and (c) Which variables by which informants are most strongly associated with high-school dropout rates after controlling for school demographics?

The authoritative model theorizes that high structure and high support have an interactive effect that produces the most positive student outcomes. Consistent with this theory, this study tested for moderation effects between structure and support measures. The present study extends previous work through use of multiple informants (students and teachers) of school climate to predict school-level dropout rates. To address these questions, this study is concerned with school-level effects and schoolwide dropout rates as distinguished from individual-level effects on students.

## Method

### Participants and Sampling

The sample of schools was obtained from the Virginia Secondary School Climate Survey, a statewide assessment of school climate and safety conditions in Virginia public secondary schools (Cornell et al., 2014). A total of 323 of 324 (97.7%) high schools participated in the survey. Eight alternative schools comprising students at risk for dropping out, and identified as being statistical outliers with respect to dropout rates, were removed from the sample. In addition, dropout rates were unavailable for one school, leaving a final analytic sample of 314 schools from which student reports were provided and 301 schools from which teacher reports were obtained. No data were missing from any of the student or teacher survey variables used in the current study.

School principals were given two options for sampling students: (a) invite all students in each Grade 9–12 to take the survey, with a goal of surveying at least 70% of all eligible students; or (b) randomly select at least 25 students from each grade to take the survey. Principals were given this flexibility so that they could choose a more or less comprehensive assessment of their students. Schools using the random sample option were provided with a random number list that corresponded to the size of their school. Written instructions explained how the random numbers should be applied to an alphabetized list of student names. Principals were advised to invite up to 50 students in each grade to take the survey to have a pool of alternates in the event that any of the first 25 selected students were unable or unwilling to participate. All students were eligible to participate except those unable to complete the survey because of limited English proficiency or an intellectual or physical disability.

Student participation rate was defined as the total number of students who participated in the survey divided by the total number invited to take the survey. The overall student participation rate was estimated to be 88.7% based on 52,012 student participants from a pool of 58,613 students asked to participate. Participation rates for schools choosing to invite all students to participate and those electing the random sampling option were 82.9% and 93.4%, respectively. According to the school principals, 82% of schools used the random sampling option.

Enrollments across the 314 schools ranged from a low of $n = 93$ to a high of $n = 4{,}072$ ($M = 1{,}201$ students). The average percentage of students eligible for free or reduced-price meals (FRPM) in the participating schools was 37.6% (range = 1.9% to 83.0%) and school enrollments were approximately 61.5% Caucasian, 22.4% African American, 8.2% Hispanic, 4.1% Asian American, and 3.8% other minority groups. These percentages align closely with state demographics obtained from the Virginia Department of Education that report the average percentage of students eligible for FRMP in these schools to be 38.0%; with approximately 60.5% Caucasian, 23.0% African American, 8.7% Hispanic, 4.1% Asian American, and 3.8% other minority groups. Sixty-one schools were in urban areas, and all other schools were located in suburbs, towns, or rural areas.

All teachers in each school were invited to participate in the survey. According to school principals, many teachers felt too burdened with other obligations to complete a voluntary survey. The teacher participation rate was 56.5%. The teachers were predominantly female (66.8%) and highly experienced, with 57.6% reporting more than 10 years of teaching, 22% reporting 6 to 10 years, 11.7% 3 to 5 years, and 8.7% 1 to 2 years. To protect teacher anonymity, no additional information about the teachers was collected.

**Measures**

School demographics and dropout rates. Enrollment (school size), percent free and reduced-price meals (FRPM), percent minority, and school location were obtained from the Virginia Department of Education. School dropout rates were obtained from the Virginia Department of Education (VDOE) for the 2009 and 2010 cohorts that graduated in 2013 and 2014, respectively. Two years were selected in order to improve the reliability of the dropout rates for each school, consistent with previous studies (Cornell, Gregory, Huang, & Fan, 2013; Lee, Cornell, Gregory, & Fan, 2011). Cohort dropout rates were calculated as the percentage of student dropouts within a longitudinal 4-year cohort group (VDOE, 2014). Average dropout rates for the 2009 and 2010 cohorts within each school served as the outcome in the current study.

**Authoritative school climate—student reports.** The student survey consisted of approximately 110 items and contained the scales for the Authoritative School Climate Survey as well as some additional items from other scales not used in this study. The median completion time was 14.4 min. Students (and teachers) were required to answer each item before proceeding to the next page of the survey; as a result, there were no missing data for survey items. The survey included two scales to measure demandingness or school structure: Disciplinary Structure and Academic Expectations, and one scale to measure responsiveness or student support, Student Support. Each scale item was answered on a 4-point Likert-scale with response options of 1= *strongly disagree*, 2 = *disagree*, 3 = *agree*, and 4 = *strongly agree*. Items for each scale were summed with higher scores indicating a more favorable school climate. School-level scores were generated by aggregating the scores for students within each school.

Disciplinary Structure was based on seven items designed to measure the perceived fairness and strictness of school discipline with items such as "The school rules are fair" and "The adults at this school are too strict."

Academic Expectations was based on five items designed to measure student perceptions of how much teachers expected of them in their academic work. Example items include "My teachers expect me to work hard." and "My teachers expect a lot from students."

Student Support was based on eight items designed to measure teacher and school staff respect and concern for students (e.g., "Most teachers and other adults at this school care about all students") and student willingness to seek help from teachers or other school staff (e.g., "There are adults at this school I could talk with if I had a personal problem").

The psychometric qualities of these scales were investigated in the context of a multilevel structural analysis of a more comprehensive authoritative school climate measurement model

(Konold & Cornell, 2015). Results from a multilevel confirmatory factor analysis of these items were favorable (CFI = .964, TLI = .959, RMSEA = .023, $SRMR_W$ = .040, and $SRMR_B$ = .096). School-level factor loadings in the form of structure coefficients for items on these scales ranged from .74 to .97 (Disciplinary Structure), .65 to .99 (Academic Expectations), and .67 to 1.0 (Student Support). School-level reliability estimates ranged from .80 to .95. In addition, support for the convergent validity of the scales was indicated through negative associations of both scales with student reports of the amount of bullying and teasing in schools and positive associations with student engagement.

**Authoritative school climate—teacher reports.** The teacher version of the authoritative school climate survey included Disciplinary Structure to measure demandingness or school structure, and Student Support to measure responsiveness or student support. Efforts to construct a teacher report measure of academic expectations were not successful because teachers tended to endorse the highest levels across all items. Each scale item was answered on a 6-point Likert-scale with response options of 1 = *strongly disagree*, 2 = *disagree*, 3 = *somewhat disagree*, 4 = *somewhat agree,* 5 = *agree*, 6 = *strongly agree*. Items were summed for each scale with higher scores indicating more favorable perceptions of school climate. School-level scores were generated by aggregating the scores for teachers within each school.

Disciplinary Structure was based on nine items designed to measure the perceived fairness and strictness of school discipline with items such as "The punishment for breaking school rules is the same for all students" and "When students are accused of doing something wrong, they get a chance to explain."

Student Support was based on 10 items designed to measure teacher and school staff respect and concern for students (e.g., "Most teachers and other adults at this school care about all students" and student willingness to seek help from teachers or other school staff (e.g., "Students feel comfortable asking for help from teachers if there is a problem with a student).

The psychometric qualities of these teacher scales were also investigated in the context of a multilevel structural analysis of a more comprehensive authoritative school climate measurement model (Huang & Cornell, 2016). Results from a multilevel confirmatory factor analysis of these items were favorable (CFI = .92, TLI = .91, RMSEA = .05, $SRMR_W$ = .04, and $SRMR_B$ = .09). School-level factor loadings in the form of structure coefficients for items on these scales ranged from .61 to 1.00 (Disciplinary Structure), and .60 to .96 (Student Support). Average school-level reliability estimates were .81 and .79, respectively. In addition, support for the convergent validity of the scales was indicated through negative associations of both scales with student reports of the amount of bullying and teasing in schools and positive associations with student engagement.

**Analytic Plan**

We expected that the three indicators of an authoritative school climate, disciplinary structure, academic expectations, and student support, would be negatively associated with school dropout rates. We examined the effects of these variables and potential moderating effects through three separate regression models. The first model used student reports of school climate, the second model used teacher reports, and the third model combined the significant student and teacher predictors obtained from models one and two. Each of these three models was examined through a two-step hierarchical regression approach. Step one included only the school-level covariates of enrollment, percent minority, percentage of students eligible for free and reduced-priced meals (FRPM), and school location. The second step added the student and/or teacher scales for authoritative school climate, including their interactions and the moderating effect of student support. Finally, in our follow-up analyses to examine interaction effects, we divided the schools into low, medium, and high groups for each of our three variables of interest: student support, academic expectations, and level of FRPM. The groups were constructed to be of equal size. For example, 1/3 of the schools with the lowest student support mean were placed in the low group, the next 1/3 of schools were in the medium group, and 1/3 of schools with the highest means were in the high group. The same process was followed to create groups based on academic expectations and on FRPM.

## Results

Before analysis, evaluation of the parametric assumption of normality revealed that dropout rates were not normally distributed across schools. A square root transformation applied to this variable resulted in more favorable distributional characteristics (e.g., skewness = .03). Collinearity was found to be an issue as indicated by variance inflation factors when student reports of disciplinary structure and support were included as separate predictors in our regression models. Consequently, these student reports were combined, through calculation of a weighted mean, into a single scale that we refer to as Disciplinary Structure and Support (DSS).

Correlations among study variables are presented in Table 1. Associations between dropout rates and all other variables were in the expected direction, statistically significant ($p < .05$), and ranged between $r = .12$ and .60. The strongest association with dropout rates was the percentage of FRPM ($r = .60$). Notably, student perception of disciplinary structure and support (DSS) was strongly associated with academic expectations, $r = .72$, $p < .01$, whereas the association for teachers was somewhat lower ($r = .32$ for support and $r = .21$ for disciplinary structure, $p < .01$). Student perception of DSS was moderately associated with teacher perceptions of structure and support ($r = .36$ and $r = .48$, respectively, $p$s < .01).

Standardized coefficients from the three regression models that varied as a function of informant type are presented in Table 2. The first step of each model included only the four school-level demographic control variables that were found to explain 38% of the variance in dropout rates ($R^2 = .38$, $p < .01$). School enrollment ($\beta = 0.17$, $p < .05$) and proportion of FRPM ($\beta = 0.69$, $p < .01$) were positively associated with dropout rates. In the second step, DSS, academic expectations, and their interactions as rated by students were added to the baseline model. In this model, student perceptions of teachers' academic expectations was found to be a significant predictor of dropout rates, $\beta = -0.19$, $p < .01$, and accounted for an additional 5% of the variance in dropout rates beyond the school-level covariates. The second step of the teacher model added disciplinary structure and student support to the baseline model that included only school-level covariates. Here, student support was found to be significantly associated with school dropout rates, $\beta = -0.16$, $p < .05$, after controlling for covariates. Teacher perceptions of their support for students explained an additional 1% of the variance in dropout rates beyond the school-level control variables.

Step two of the final analysis was developed by carrying forward variables that were statistically significant in the student and teacher informant models (see last column of Table 2). Results indicated that student perceptions of teachers' academic expectations, $\beta = -0.10$, $p < .01$, teacher perceptions of their support for students, $\beta = -0.17$, $p < .05$, and the interaction between the two, $\beta = -0.17$, $p < .01$, were significantly associated with school dropout rates. In combination, the school climate predictors accounted for an additional 7.5% of the variance in school dropout rates beyond the school-level covariates.

To further investigate the moderating role of student support on teacher academic expectations and school dropout rates, schools were equally divided into three groups (low/medium/ high) for both student support and academic expectations, with dropout rates adjusted for the influence of school demographics. Inspection of Figure 1 reveals that dropout rates were lowest in schools characterized by high student support and high academic expectations, whereas the level of academic expectations did not show a material difference in adjusted dropout rates when student support was low. Notably, in high support schools the dropout rate was 5.7% in schools with low academic expectations, but 3.0% in schools with high academic expectations, a difference of ([5.7–3.0]/5.7 = 47.4%). In contrast, in low support schools there was little difference in dropout rates associated with academic expectations; that is, dropout rates were 6.0%, 7.1%, and 6.5%, respectively, in schools with low, medium, and high academic expectations.

Figures 2 and 3 illustrate the role of school levels of free and reduced priced meals on the relations between student support (see Figure 2) and academic expectations (see Figure 3) on model adjusted dropout rates. The FRPM rates are 16.6% in the low group (schools with few low-income students), 37.2% in the medium group, and 58.8% in the high group (schools with the most low-income students). In both figures it can be seen that schools characterized by higher percentages of students receiving FRPMs tend to have higher dropout rates, with a downward trend across schools characterized by higher levels of student support and higher academic expectations.

Notably, Figure 2 shows that, in schools with high levels of FRPM, schools with low student support have a dropout rate of 7.9% whereas schools with high student support have a dropout rate of 6.4%, a difference of 19.0%. For schools with low levels of FRPM, schools with low student support have a dropout rate of 3.8% whereas schools with high student support have a dropout rate of 2.2%, a difference of 42.1%.

Figure 3 shows that, in schools with high levels of FRPM, schools with low academic expectations have a dropout rate of 7.7% whereas schools with high academic expectations have a dropout rate of 6.5%, a difference of 15.6%. For schools with low levels of FRPM, schools with low academic expectations have a dropout rate of 3.8% whereas schools with high academic expectations have a dropout rate of 2.4%, a difference of 36.8%.

## Discussion

The present study found that an authoritative school climate was associated with lower dropout rates in a large statewide sample of high schools. In the final analyses, study measures explained 45.5% of the variance in dropout rates, with school demographics accounting for 38.9% and features of an authoritative school climate accounting for 7.5% of the variance in school dropout rates. School-level findings are especially important in light of the sustained national effort to improve poorly performing schools and reduce dropout rates (Legters & Balfanz, 2010; National Dropout Prevention Center/Network, 2015). As Legters and Balfanz (2010, p. 17) observed, “Because students drop out of schools, not districts, communities, or states, ending the dropout crisis must include large-scale transformation of the schools in which large percentages of students are falling off the graduation path.”

Many authorities have concluded that high academic expectations are critical to encouraging student achievement and lowering the dropout rates (Legters & Balfanz, 2010; Werblow, Urick, & Duesbery, 2013). When teachers have high expectations of their students, students become more engaged, feel more competent, and perform better (Feldlaufer, Midgley, & Eccles, 1988; Lee & Loeb, 2000; Midgley, Feldlaufer, & Eccles, 1988; Stipek & Daniels, 1988), which makes them less likely to drop out of school. Conversely, when teachers have low expectations and doubt their ability to succeed, students may develop negative self-perceptions that interfere with their ability to learn (Rosenthal & Jacobson, 1968). A pattern of low expectations and low engagement can lead to repeated academic failure, alienation, and eventual dropout from school (Janosz, Archambault, Morizot, & Pagani, 2008; Rumberger, 2011).

Our findings suggest that high academic expectations in isolation may not be sufficient, but may interact with other qualities of the school climate. In particular, the relation between high academic expectations and lower dropout rates was moderated by supportive student-teacher relationships. In schools with high levels of student support, the dropout rate was approximately 34.8% lower than in schools with low levels of student support. Consistent with authoritative school climate theory, students may be more responsive to teacher encouragement for them to work hard in their studies when their teachers are supportive and concerned (Gregory & Cornell, 2009).

Schools with high numbers of students from low-income families have the highest dropout rates and are often the target for school improvement efforts (Balfanz & Legters, 2004; Legters & Balfanz, 2010; Rumberger, 2011). In the present study, schools with the highest percentage of students with FRPM status (school $M$ = 58.8%) had a dropout rate of 7.4% whereas schools with the lowest FRPM (school $M$ = 16.6%) had a dropout rate of 3.1%. Although educators cannot change the poverty level of their students, they may be able to ameliorate its negative impact through a school climate that encourages and promotes academic achievement. Among the 105 schools with the highest FRPM, the dropout rate was 19.0% lower in schools with high support than in schools with low support. Similarly, in the high FRPM schools, the dropout rate was 15.6% lower in schools where the students reported high academic expectations by their teachers than in schools with low academic expectations. The variability among schools with high numbers of low-income students suggests that there is potential to reduce the negative impact of poverty on high school completion.

A recent investigation used a multilevel multitrait-multimethod structural model using the same student support scale as the present study to examine the degree to which ratings obtained by middle school students and teachers were influenced by the targeted trait versus extraneous sources that could be attributed to the rater (Konold & Cornell, 2015). This study found that ratings of school climate obtained from both students and teachers demonstrated high levels of convergent validity at the school level, and that both student and teacher perceptions of student support were associated with higher student engagement and lower rates of peer aggression at the school level. Informed by these results, the present study examined how student and teacher perceptions of school climate could be combined to provide a more comprehensive assessment that included measures of structure, academic expectations, and school-level covariates that were not considered in the Konold and Cornell (2015) investigation. In the present study, the initial analyses produced different results for students and teachers. Student perceptions that teachers had high academic expectations were associated with lower dropout rates, while teacher perceptions that students felt supported were also associated with lower dropout rates. The statistically significant predictors for students and teachers were combined in the final analysis, which found both direct and interaction effects. As expected, both student support (as reported by teachers) and high academic expectations (as reported by students), and the interaction of these two measures, were associated with dropout rates.

In previous studies using three independent samples, structure and support scales were highly correlated, but could be statistically distinguished in factor analyses (Gregory et al., 2010; Konold et al., 2014; Konold & Cornell, 2015) and demonstrated evidence of criterion-related validity through their association with student outcomes such as peer aggression and aggression toward teachers (Cornell et al., 2015; Gregory, Cornell, & Fan, 2012; Gregory et al., 2010). However, in the present study the student scales for disciplinary structure and student support demonstrated evidence of collinearity in our regression models and were combined into a single variable, termed disciplinary structure and support (DSS). This problem did not present with the teacher scales for disciplinary structure and student support and so they were retained as separate variables. One direction for future study is to assess whether the constructs of disciplinary structure, which means that students perceive school discipline to be strict, but fair, can be adequately distinguished from student support, which means that students feel that their teachers have respect and concern for them. Although the two scales seem conceptually distinct, students may regard teacher fairness in discipline and teacher concern as so closely related that it is difficult to discriminate between them.

### Limitations and Directions for Future Research

Correlational findings cannot establish causal relationships and are open to multiple interpretations. There may be other variables contributing to the relations between school climate and dropout rates. For example, one study reported that the prevalence of teasing and bullying was predictive of high school dropout rates (Cornell et al., 2013), and this could be a mediating variable in the association of school climate and dropout.

These findings were obtained in a large and diverse sample of schools representing almost all of the state's public high schools. However, the sample was limited because the schools were from a single state and the teacher survey had a participation rate of 57%. Continued research in other states and with greater teacher participation is needed.

### Practice Implications

School psychologists may find that many widely used school improvement programs and strategies have a positive impact on school climate that is consistent with the authoritative school climate model. For example, Positive Behavioral Interventions and Supports (PBIS) is a school-wide approach to discipline designed to create safe, predictable, and positive school environments (Bradshaw, Waasdorp, & Leaf, 2012). Under the PBIS model, schools establish school-wide expectations for student behavior that stress positive goals and establish a reward system to reinforce positive behavior. Sprague and Horner (2006) identified a number of key practices in PBIS: clearly defining behavioral expectations, proactively teaching what those expected behaviors look like in various school settings, rewarding students for compliance with behavioral expectations, and providing clear and fair consequences for behavioral violations. This approach seems to dovetail with the basic principles of an authoritative school climate and might be used more explicitly to inform teacher behavior and enhance the overall effectiveness of the model. A PBIS model will likely be most effective when students perceive that the disciplinary system is fair and that positive reinforcement efforts reflect respect for students.

Another example of a school improvement program compatible with authoritative school climate theory is My Teaching Partner-Secondary (MTP-S), which provides coaching for teachers on ways to improve their interactions with students (Gregory, Allen, Mikami, Hafen, & Pianta, 2014). A primary goal of MTP-S is to build an emotionally supportive relationship between teachers and students, which seems to parallel the authoritative conception of support. According to the MTP-S model, such relationships are characterized by feelings of warmth and connection, as well as responsiveness to the student's academic and social/emotional needs. School psychologists may want to apply similar authoritative principles in their consultation with teachers. The message for teachers is that they will find that high academic expectations for their students will be most effective when the students perceive that their teachers have respect and concern for them. Positive statements and expressions of interest in the classroom and in informal interactions may help strengthen the student's perceptions of his or her teacher and improve his or her classroom engagement.

## Conclusion

The improvement of school climate has become a nationwide goal because of its recognized impact on school quality and student outcomes (Dary & Pickeral, 2013; Thapa et al., 2013). However, there is little consensus on the key qualities of a positive school climate. The authoritative school climate model provides a conceptual framework that can more precisely identify key elements of school climate associated with positive student outcomes.

This study found support for the view that schools must promote high academic expectations for their students to reduce the dropout problem, but with an important qualification. Consistent with authoritative school climate theory, high academic expectations had their strongest association with school completion when teacher-student relationships were supportive. In the absence of a supportive school climate, high academic expectations may not have the desired positive impact on student performance. Both high academic expectations and supportive student–teacher relationships were associated with lower dropout rates even in schools with a high proportion of low-income students. Studies of school improvement efforts would be the best way to demonstrate the combined effects of school climate change that increases both academic expectations and the quality of teacher-student relationships.

## References


American Institutes for Research. (2013). School climate survey compendia. Retrieved from http:// safesupportivelearning.ed.gov/topic-research/schoolclimate-measurement/school-climatesurveycompendium

Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis: Which high schools produce the nation's dropouts? Where are they located? Who attends them?* Retrieved from http://files.eric.ed .gov/fulltext/ED484525.pdf

Baumrind, D. (1968). Authoritarian vs. authoritative parental control. *Adolescence, 3,* 255–272.

Belfield, C., & Levin, H. M. (Eds.). (2007). *The price we pay: Economic and social consequences of inadequate education*. Washington, DC: Brookings Institution Press.

Berg, J. K., & Cornell, D. (2016). Authoritative school climate, aggression toward teachers, and teacher distress in middle school. *School Psychology Quarterly, 31*(1), 122-139. http://dx.doi.org/10.1037/spq0000132

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130,* e1136–e1145. http://dx.doi .org/10.1542/peds.2012-0243

Brault, M. C., Janosz, M., & Archambault, I. (2014). Effects of school composition and school climate on teacher expectations of students: A multilevel analysis. *Teaching and Teacher Education*, *44,* 148–159. http://dx.doi.org/10.1016/j.tate.2014.08 .008

Bridgeland, J. M., Dilulio, J. J., & Balfanz, R. (2009). *On the front lines of schools: Perspectives of teachers and principals on the high school dropout problem*. Washington, DC: Civic Enterprises. Retrieved from http://www.civicenterprises.net/pdfs/ frontlines.pdf

Bridgeland, J. M., Dilulio, J. J., & Morison, K. B. (2006). *The silent epidemic: Perspectives of high school dropouts*. Washington, DC: Civic Enterprises. Retrieved from http://www.civicenterprises.net/pdfs/ thesilentepidemic3-06.pdf

Christle, C. A., Jolivette, K., & Nelson, C. M. (2007). School characteristics related to high school dropout rates. *Remedial and Special Education, 28,* 325–339. http://dx.doi.org/10.1177/ 07419325070280060201

Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record, 111*, 180–213.

Cornell, D., Gregory, A., Huang, F., & Fan, X. (2013). Perceived prevalence of teasing and bullying predicts high school dropout rates. *Journal of Educational Psychology, 105,* 138–149. http://dx .doi.org/10.1037/a0030416

Cornell, D., Huang, F., Konold, T., Meyer, P., Lacey, A., Nekvasil, E., . . . Shukla, K. (2014). *School climate and safety in Virginia high schools: Tech. Rep. No. of the Virginia Secondary School Climate Survey, 2014: Results for 9th-12th grade students and teachers*. Charlottesville, VA: Curry School of Education, University of Virginia.

Cornell, D. G., & Mayer, M. J. (2010). Why do school order and safety matter? *Educational Researcher, 39,* 7–15. http://dx.doi.org/10.3102/ 0013189X09357616

Cornell, D., Shukla, K., & Konold, T. R. (2015). Peer victimization and authoritative school climate: A multilevel multivariate approach. *Journal of Educational Psychology*, 107(4), 1186-1201. http://dx.doi.org/10.1037/edu0000038

Dary, T., & Pickeral, T. (Eds.). (2013). *School climate: Practices for implementation and sustainability. A school climate practice brief, Number 1*. New York, NY: National School Climate Center.

Feldlaufer, H., Midgley, C., & Eccles, J. S. (1988). Student, teacher, and observer perceptions of the classroom environment before and after the transition to junior high school. *The Journal of Early Adolescence, 8,* 133–156. http://dx.doi.org/10 .1177/0272431688082003

Freeman, J., & Simonsen, B. (2015). Examining the impact of policy and practice interventions on high school dropout and school completion rates: A systematic review of the literature. *Review of Educational Research, 85*, 205–248. https://doi.org/10.3102/0034654314554431

Goddard, R. D., Sweetland, S. R., & Hoy, W. K. (2000). Academic emphasis of urban elementary schools and student achievement in reading and mathematics: A multilevel analysis. *Educational Administration Quarterly, 36,* 683–702. http://dx .doi.org/10.1177/00131610021969164

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2014). Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in the Schools, 51,* 143–163. http://dx .doi.org/10.1002/pits.21741

Gregory, A., & Cornell, D. (2009). "Tolerating" adolescent needs: Moving beyond zero tolerance policies in high school. *Theory into Practice, 48,* 106–113. http://dx.doi.org/10.1080/00405840902776327

Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for Black and White high school students. *American Educational Research Journal, 48,* 904–934. http://dx.doi.org/ 10.3102/0002831211398531

Gregory, A., Cornell, D., & Fan, X. (2012). Teacher safety and authoritative school climate in high schools. *American Journal of Education, 118*, 401–425. http://dx.doi.org/10.1086/666362

Gregory, A., Cornell, D., Fan, X., Sheras, P., Shih, T. H., & Huang, F. (2010). Authoritative school discipline: High school practices associated with lower bullying and victimization. *Journal of Educational Psychology, 102,* 483–496. http://dx.doi .org/10.1037/a0018562

Huang, F., & Cornell, D. (2016). Multilevel factor structure, concurrent validity, and test-retest reliability of the high school teacher version of the Authoritative School Climate Survey. *Journal of Psychoeducational Assessment, 34*, 536-549. https://doi.org/10.1177/0734282915621439

Huang, F. L., Cornell, D. G., Konold, T., Meyer, J. P., Lacey, A., Nekvasil, E. K., . . . Shukla, K. D. (2015). Multilevel factor structure and concurrent validity of the teacher version of the Authoritative School Climate Survey. *The Journal of School Health, 85,* 843–851. http://dx.doi.org/10.1111/ josh.12340

Janosz, M., Archambault, I., Morizot, J., & Pagani, L. S. (2008). School engagement trajectories and their differential predictive relations. *Journal of Social Issues, 64,* 21–40. http://dx.doi.org/10 .1111/j.1540-4560.2008.00546.x

Johnson, S. L. (2009). Improving the school environment to reduce school violence: A review of the literature. *The Journal of School Health, 79*, 451– 465. http://dx.doi.org/10.1111/j.1746-1561.2009 .00435.x

Kearney, W. S., & Peters, S. (2013). A comparison of teacher and student perceptions of elementary classroom climate. *National Forum of Educational Administration & Supervision Journal, 31,* 20–37.

Konold, T., & Cornell, D. (2015). Measurement and structural relations of an Authoritative School Climate model: A multi-level latent variable investigation. *Journal of School Psychology*, *53*, 447-461. https://doi.org/10.1016/j.jsp.2015.09.001

Konold, T., & Cornell, D. (2015). Multilevel multitrait-multimethod latent analysis of structurally different and interchangeable raters of school climate. *Psychological Assessment, 27,* 1097–1109. http://dx.doi.org/10.1037/pas0000098

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. http://dx.doi.org/10.1037/ spq0000062

Larzelere, R. E., Morris, A. S. E., & Harrist, A. W. (Eds.). (2013). *Authoritative parenting: Synthesizing nurturance and discipline for optimal child development*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/ 13948-000

Lee, J. S. (2012). The effects of the teacher-student relationship and academic press on student engagement and academic performance. *International Journal of Educational Research, 53,* 330– 340. http://dx.doi.org/10.1016/j.ijer.2012.04.006

Lee, T., Cornell, D., Gregory, A., & Fan, X. (2011). High suspension schools and dropout rates for Black and White students. *Education & Treatment of Children, 34,* 167–192. http://dx.doi.org/10 .1353/etc.2011.0014

Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. *American Educational Research Journal, 40,* 353–393. http://dx.doi.org/10.3102/ 00028312040002353

Lee, V. E., & Loeb, S. (2000). School size in Chicago elementary schools: Effects on teachers' attitudes and students' achievements. *American Educational Research Journal, 37,* 3–31. http://dx.doi .org/10.3102/00028312037001003

Lee, V. E., & Smith, J. B. (1999). Social support and achievement for young adolescents in Chicago: The role of school academic press. *American Educational Research Journal, 36,* 907–945. http:// dx.doi.org/10.3102/00028312036004907

Legters, N., & Balfanz, R. (2010). Do we have what it takes to put all students on the graduation path? *New Directions for Youth Development, 2010,* 11– 24. http://dx.doi.org/10.1002/yd.359

Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research, 79,* 464–490. http://dx.doi.org/10.3102/ 0034654308326158

Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement, 16,* 249–279. http://dx.doi.org/10 .1080/09243450500114884

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1988). The transition to junior high school: Beliefs of preand posttransition teachers. *Journal of Youth and Adolescence, 17,* 543–562. http://dx.doi.org/10 .1007/BF01537831

Mitchell, M. M., Bradshaw, C. P., & Leaf, P. J. (2010). Student and teacher perceptions of school climate: A multilevel exploration of patterns of discrepancy. *The Journal of School Health, 80,* 271–279. http://dx.doi.org/10.1111/j.1746-1561 .2010.00501.x

National Center for Education Statistics. (2014). *Digest of education statistics*. Retrieved from https:// nces.ed.gov/programs/digest/d14/tables/dt14_ 219.70.asp

National Center for Education Statistics. (2015). *Trends in high school dropout and completion rates in the United States: 1972–2012*. Retrieved from http://nces.ed.gov/pubs2015/2015015.pdf

National Dropout Prevention Center/Network. (2015). *Effective strategies*. Retrieved from http:// dropoutprevention.org/effective-strategies/

Pellerin, L. A. (2005). Applying Baumrind's parenting typology to high schools: Toward a middlerange theory of authoritative socialization. *Social Science Research, 34,* 283–303. http://dx.doi.org/ 10.1016/j.ssresearch.2004.02.003

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York, NY: Holt, Rinehart & Winston.

Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge, MA: Harvard University Press. http://dx.doi.org/10.4159/harvard .9780674063167

Sprague, J. R., & Horner, R. H. (2006). Schoolwide positive behavioral supports. In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), *Handbook of school violence and school safety: From research to practice* (pp. 413–427). New York, NY: Routledge.

Stetser, M. C., & Stillwell, R. (2014). *Public high school four-year on-time graduation rates and event dropout rates: School years 2010–11 and 2011–12*. First look. NCES 2014–391. Washington, DC: National Center for Education Statistics.

Stipek, D., & Daniels, D. (1988). Declining perceptions of competence: A consequence of changes in the child or in the educational environment? *Journal of Educational Psychology, 80,* 352–356.http://dx.doi.org/10.1037/0022-0663.80.3.352

Suh, S., & Suh, J. (2007). Risk factors and levels of risk for high school dropouts. *Professional School Counseling, 10,* 297–306. http://dx.doi.org/10 .5330/prsc.10.3.w26024vvw6541gv7

Swanson, C. B. (2009). *Cities in crisis 2009: Closing the graduation gap: Educational and economic conditions in America's largest cities*. Bethesda, MD: Editorial Projects in Education, Inc.

Thapa, A., Cohen, J., Guffey, S., & HigginsD'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83,* 357–385. http://dx.doi.org/10.3102/ 0034654313483907

Virginia Department of Education. (2014). *Graduation, completion, dropout & postsecondary data*. Retrieved from http://doe.virginia.gov/statistics_ reports/graduation_completion/index.shtml

Wang, M. T., & Degol, J. L. (2015). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*. Advance online publication. http://dx.doi .org/10.1007/s10648-015-9319-1

Werblow, J., & Duesbery, L. (2009). The impact of high school size on math achievement and dropout rate. *The High School Journal, 92,* 14–23. http:// dx.doi.org/10.1353/hsj.0.0022

Werblow, J., Urick, A., & Duesbery, L. (2013). On the wrong track: How tracking is associated with dropping out of high school. *Equity & Excellence in Education, 46,* 270–284. http://dx.doi.org/10 .1080/10665684.2013.779168

Table 1

*Correlations Among Study Variables and Descriptive Statistics*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Enrollment | — | .40 | .41 | .16 | .11 | .16 | .10 | .24 | .12 |
| 2. % Minority | | — | .37 | .46 | .21 | .35 | .32 | .20 | .33 |
| 3. % FRPM | | | — | .29 | .33 | .17 | .29 | .17 | .60 |
| 4. Urbanicity | | | | — | .10 | .21 | .17 | .15 | .15 |
| 5. DSS | | | | | — | .36 | .48 | .72 | .36 |
| 6. Structure: Teacher | | | | | | — | .74 | .21 | .23 |
| 7. Support: Teacher | | | | | | | — | .32 | .32 |
| 8. Academic expectations | | | | | | | | — | .27 |
| 9. Dropout rates | | | | | | | | | — |
| Mean | 1200.95 | .39 | .38 | .19 | 2.85 | 4.46 | 5.45 | 3.16 | 5.28 |
| *SD* | 705.29 | .26 | .19 | .40 | .15 | .32 | .24 | .10 | 3.23 |
| Skew | .72 | .40 | .12 | 1.56 | .28 | .21 | .58 | .51 | .67 |
| Kurtosis | .62 | .80 | .80 | .43 | .54 | .89 | 4.18 | 1.35 | .35 |

Note. Enrollment = total number of students enrolled in a school; DSS = combined student reports of Disciplinary Structure and Support; *SD* = standard deviation.

* $p < .05$, ** $p < .01$.

Table 2
*Standardized Regression Coefficients*

| Predictors | Student $\beta_{Step1}$ | Student $\beta_{Step2}$ | Teacher $\beta_{Step1}$ | Teacher $\beta_{Step2}$ | Student and teacher $\beta_{Step1}$ | Student and teacher $\beta_{Step2}$ |
|---|---|---|---|---|---|---|
| Enrollment | **.167**** | **.168**** | **.167**** | **.152*** | **.167**** | **.150**** |
| *% Minority | .007 | .048 | .007 | -.025 | .007 | — |
| % FRPM | **.694**** | **.634**** | **.694**** | **.655**** | **.694**** | **.628**** |
| Urbanicity | -.095 | -.072 | -.095 | **-.107*** | -.095 | -.064 |
| $DSS_{Student}$ | — | -.031 | — | — | — | — |
| EXP | — | **-.193**** | — | — | — | **-.102**** |
| $STR_{Teacher}$ | — | — | — | -.016 | — | — |
| $SUP_{Teacher}$ | — | — | — | **-.155*** | — | **-.174*** |
| EXP* $DSS_{Student}$ | — | -.079 | — | — | — | — |
| STR* $SUP_{Teacher}$ | — | — | — | -.037 | — | — |
| EXP* $SUP_{Teacher}$ | — | — | — | — | — | **-.172**** |
| $R^2$ | **.380**** | **.430**** | **.380**** | **.389**** | **.380**** | **.455**** |
| $R^2$ | — | **.050**** | — | **.009**** | — | **.075**** |

*Note*. DSS = combined student reports of Disciplinary Structure and Support; EXP = student perceived academic expectations; STR = Disciplinary Structure; SUP = Student Support.
* $p < .05$ ** $p < .01$.

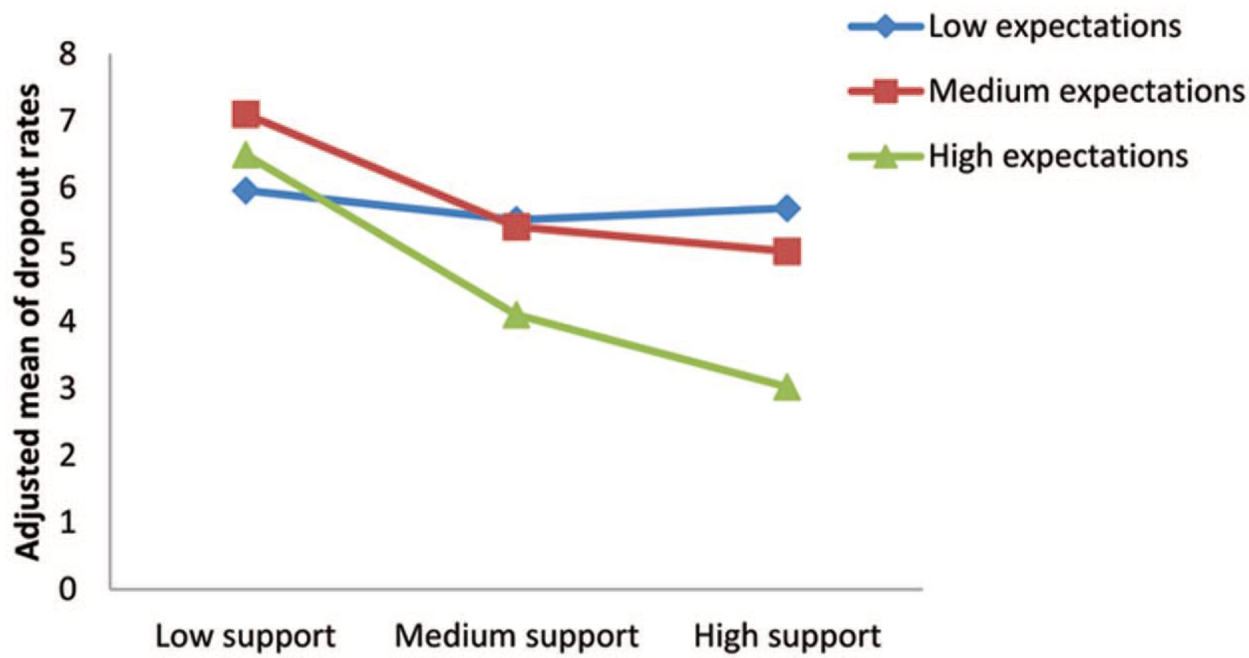


*Figure 1.* Moderating effect of student perceptions of teachers' academic expectations on the relations between student support and model-adjusted dropout rate means.

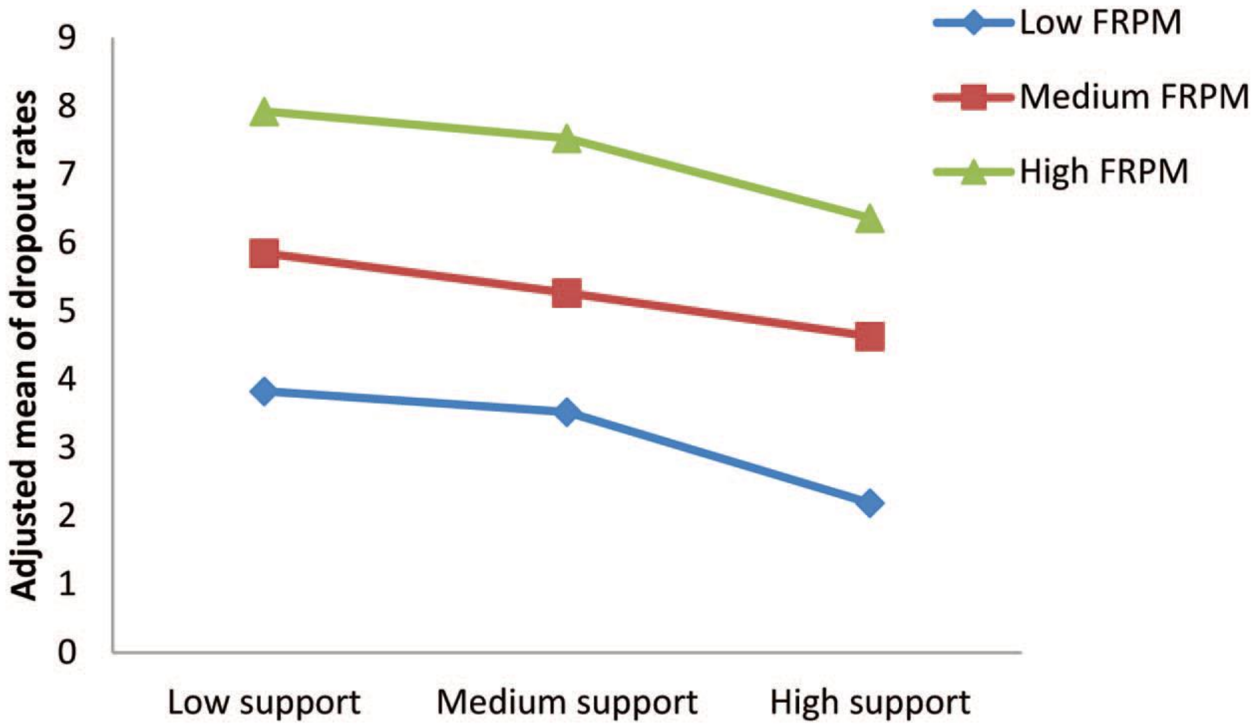


*Figure 2.* Model-adjusted dropout rate means as a function of student support and school FRPM percentages.

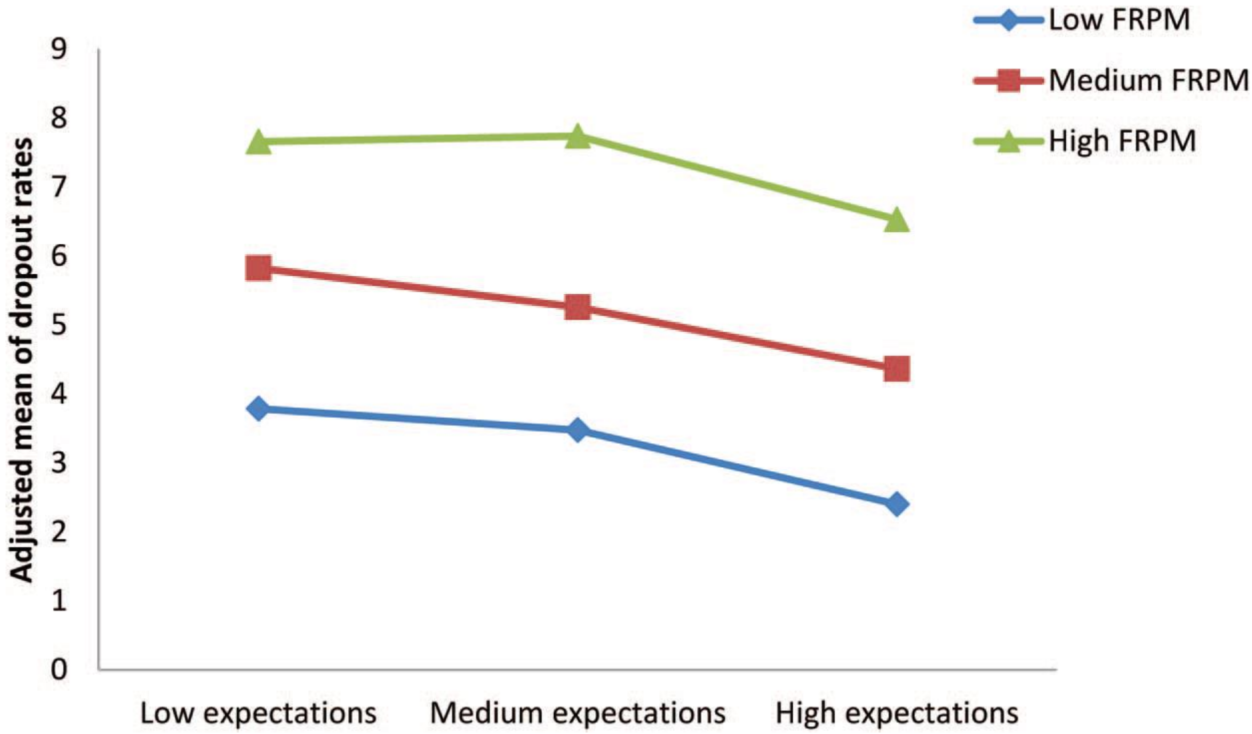


*Figure 3.* Model-adjusted dropout rate means as a function of student perceptions of teachers' academic expectations and school FRPM percentages.

## Manuscript Three

# Moving to the Next Level: Doubly Latent Multilevel Mediation Models with School Climate Illustrations

Yuane Jia & Timothy Konold

Curry School of Education, University of Virginia

Author Note

Yuane Jia and Timothy R. Konold, Department of Educational Leadership, Foundations, and Policy, Curry School of Education, University of Virginia.

We thank members of the project research team including Anna Grace Burnette, Dewey Cornell, Brittany Crowley, Katrina Debnam, Francis Huang, Jennifer Maeng, Marisa Malone, Patrick Meyer, and Shelby Stohlman. This project was supported by Grant #2012-JF-FX-0062 awarded by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice; and Grant #NIJ 2014-CK-BX-0004 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect those of the Department of Justice, the funding agency.

Correspondence concerning this article should be addressed to 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

Abstract

Traditional observed variable multilevel models for evaluating indirect effects are limited by their inability to quantify measurement and sampling error that can result from the sampling of observed variables and persons within level two units, respectively. They are further restricted by being unable to fully separate within- and between-level effects without bias. Doubly latent multilevel mediation models reduce these biases by decomposing the observed within-level indicators into within- and between-level latent components. This decomposition simultaneously controls for both measurement error at each level and sampling error in the aggregation of individual values that serve as indicators of higher-level constructs. We illustrate the usefulness of this approach in investigating the mediating role of engagement in associations between reports of school climate and a variety of student outcomes (i.e., self-report GPA, PTB, suspension rates). Analyses are e based on a statewide high school sample of 59,581 students and 11,336 teachers from 296 schools. Results reveal that student engagement serves  as a mediator between school climate factors and student outcomes.

*Keywords*: doubly latent, multilevel mediation, school climate, student engagement, student outcomes

Moving to the Next Level: Doubly Latent Multilevel Mediation Models with School Climate Illustrations

Organizational research typically requires special analytic attention to model multilevel data that arise from the sampling of observations that reside within higher-order structures. This need emerges in a variety of disciplines and substantive areas of inquiry where individual level one (L1) informants are used to obtain measurements of level two (L2) group constructs. Examples include research on neighborhood safety where parents and youth report on characteristics of their surroundings (Luo, Breslau, Gardiner, Chen, & Breslau, 2014), workplace leadership where managers' report on the culture of their organizations (Dyer, Hanges, & Hall, 2005), and school climate where students and teachers report on the features of their schools (Konold & Shukla, 2017).

In the sections below, we describe several measurement issues that can arise when estimating organizational constructs, discuss potential methodological problems with the use of traditional multilevel models when L2 organizational constructs are measured through the use of sampled L1 informants, and point to recent developments in multilevel modeling techniques that can be useful in overcoming these issues when testing direct and indirect effects. This approach, known as doubly latent multilevel structural equation modeling (MSEM; Lüdtke et al., 2011; Marsh et al., 2009), has not yet been widely adopted in the substantive literature despite its adaptability to a variety of research designs. We illustrate the application of this approach to school climate research given continued interest in providing students with healthy learning environments and its importance in national policy (e.g., 2015 Every Student Succeeds Act; U.S. Department of Education, 2017). In doing so, we describe the interpretable elements of the model, and provide annotated Mplus syntax that users can adapt for their own work.

**Measurement Issues in Estimating Organizational Constructs**

Higher-order organizational constructs like school climates are typically estimated through the collection of the informant (e.g., students and/or teachers) perspectives that are sampled from within schools. However, more than individual experiences, higher-order organizational constructs reflect group-level characteristics that are indicative of collective experiences, shared beliefs, values, and attitudes among the members within each group (Cohen, McCabe, Michelli, & Pickeral, 2009). Despite early recognition that "studying individuals as perceivers within the classrooms could be interesting, but [it] is a problem quite separate from the measurement of environments" (Cronbach, 1976, p. 18), many researchers continue to use ratings from a single L1 informant as a proxy for an L2 group-level construct, or obtain group-level estimates through aggregation of individual observed variable ratings within groups (e.g., averaging individual observed ratings across informants in each school to obtain a school-level measure). These practices give rise to several noteworthy methodological issues.

First, in the context of school climate research, the referent of L1 informant ratings is typically the school and not the individuals' themselves. These perceptions can vary as a function of the peer groups that the informants' interact with, and the unique roles that students and teachers play within the school. Consequently, individual observed variable ratings contain sources of variance that are shared with other school informants and are representative of true L2 trait variance, and residual variance sources that are unique and unrelated to the L2 trait being measured (Marsh et al., 2012). Aggregation of individual observed variable ratings conflates these sources of variance through inclusion of non-shared residual variance components that are of no substantive value in estimating school-level traits (Cronbach, 1976; Marsh et al., 2012;

Van Mierlo, Vermunt, & Rutte, 2009). As a result, average observed variable ratings across informants within a school may not appropriately represent either the L2 construct or individual differences in perceptions of climate (Morin, Marsh, Nagengast, & Scalas, 2014) as failure to disaggregate residual variance components in L2 measures may introduce bias in estimates of group-level effects (Lüdtke et al., 2008; Shin & Raudenbush, 2010).

Second, the issue of sampling error arises when individual responses are averaged within L2 units to compute an observed group-mean that is then used to represent L2 constructs (Nagengast & Marsh, 2011). Because school climate is typically assessed by obtaining ratings from a sample of individual informants (e.g., students or teachers) from within the broader school population, different random samples may produce different estimates of the schools' population mean. The sampling error, or unreliability, inherent in a sample estimate like the mean can lead to bias in estimating substantive associations between L2 school climate constructs and other L2 outcomes (Shin & Raudenbush, 2010). The effect becomes increasingly pronounced as both the relative and absolute size of the L1 informant units vary across a large number of L2 school units. The relative size of the randomly sampled informants, from a potential population of students or teachers within schools, impacts the size of the sampling distribution for a statistic like the school mean. All things equal, larger sample sizes produce narrower sampling distributions and smaller standard errors. Moreover, the absolute size of the sample can play a role when sampling an unequal number of informants across different schools. Here, greater weight is given to schools with a larger number of informants when ratings are aggregated across schools (Wang & Degol, 2016). According to Marsh et al. (2012), sampling error in L2 constructs obtained through aggregation of L1 response ratings is "a function of the average agreement among individuals in the same group and the number of sampled individuals in each group" (p.111). When there is strong agreement on a large number of items among informants within the same school, estimates of sampling error will be smaller. Consequently, these L1 responses will provide better estimates of the L2 school trait being measured.

Third, issues pertaining to the sampling of informants within schools can also be extended to the sampling of items used to create a measurement scale, when items are drawn from a population of potential items (Marsh et al., 2009). Although most organizational constructs are measured through the use of multiple indicators, and their dimensionality has been substantiated through the use of exploratory and/or confirmatory factor analytic procedures, operational calculation of organizational scores on these scales typically takes the form of obtaining scale means across items or summing the items to obtain a total score. These approaches implicitly assume that the items are measured without error (i.e., perfectly reliable) and that each item contributes equally to the measurement of a given trait (i.e., unit weighting; DiStefano, Zhu, & Mindrila, 2009). In addition, measurement error and reliability are in part a function of the number of items on a scale, and their associations with one another. As the number of scale items and their correlations increase, higher reliabilities can be achieved (Schmitt, 1996). Conversely, smaller numbers of items with lower associations among them can result in less reliable scales that increase the bias when estimating their associations with other observed variable scales (Macmann, Barnett, & Lopez, 1993).

**Methodological Issues in Observed Variable Multilevel Mediation Analysis**

Mediation analysis focuses on the processes by which variables are related to one another, and has been widely used to understand the mechanisms of how one variable exerts its effects on others. In the context of multilevel data, mediating effects can be modeled at the individual and/or the cluster level, and the mechanisms of mediation may not necessarily be the same across levels (Hofmann & Gavin, 1998). When the predictor variable X, mediator variable M, and outcome variable Y are assessed in a multilevel context, both intercepts and slopes can vary randomly across clusters. Following Kenny et al. (2003) and Preacher (2015), evaluations of multilevel mediation in a traditional multilevel modeling (MLM) framework can be written as follows when X, M, and Y are all assessed at level 1 (L1):

Level 1

$$Y_{ij} = \beta_{Y0j} + \beta_{YXj}X_{ij} + \beta_{YMj}M_{ij} + \varepsilon_{Yij} \quad \text{[Equation Set 1]}$$
$$M_{ij} = \beta_{M0j} + \beta_{MXj}X_{ij} + \varepsilon_{Mij}$$

Level 2

$$\beta_{Y0j} = \gamma_{Y0} + u_{Y0j} \quad \text{[Equation Set 2]}$$
$$\beta_{YXj} = \gamma_{YX0} + u_{YXj}$$
$$\beta_{YMj} = \gamma_{YM0} + u_{YMj}$$
$$\beta_{M0j} = \gamma_{M0} + u_{M0j}$$
$$\beta_{MXj} = \gamma_{MX0} + u_{MXj}$$

Where, i and j denote individuals and clusters, respectively; Y = the outcome variable measured at Level 1, X = the predictor variable measured at L1, M = the mediating variable measured at L1, $\varepsilon_{ij}$ = L1 residuals, $u_j$ = L2 residuals, $\beta_{MXj}$ = the random effect of X on M, $\beta_{YMj}$ = the random effect of M on Y conditional on X, $\beta_{YXj}$ = the random effect of X on Y conditional on M, $\beta_{Y0j}$ = the random intercept of Y regressed on X and M, $\beta_{M0j}$ = the random intercept of M regressed on X, $\gamma_{YX0}$ = the average effect of X on Y conditional on M, $\gamma_{YM0}$ = the average effect of M on Y conditional on X, $\gamma_{MX0}$ = the average effect of X on M, $\gamma_{Y0}$ = the average intercept of Y regressed on X and M across all L2 units (i.e., grand mean of Y), and $\gamma_{M0}$ = the average intercept of M regressed on X across all L2 units (i.e., grand mean of M).

As described by Kenny et al. (2003), when the effect of X on M ($\beta_{MXj}$) and the effect of M on Y ( $\beta_{YMj}$) conditional on X are random (i.e., not necessarily independent), the indirect effect is no longer the simple product of the two fixed slopes ($\gamma_{MX0} * \gamma_{YM0}$) as would be the case in a single level model. Rather, an additional term is introduced pertaining to the covariance between the random effects that should be taken into consideration when estimating the indirect effect and its associated standard error. In this observed variable framework, Bauer et al. (2006)

illustrated how the estimation of these random effects could be obtained by combining the two L1 equations and mediating effects simultaneously.

Despite the advantages of traditional multilevel models (MLM) over single level models in general (e.g., more accurate estimation of standard errors when the independence assumption is violated), three limitations of these observed variable approaches have been identified. First, MLM approaches use observed scores that do not control for issues pertaining to the sampling of individuals or items (Marsh et al., 2012), which is more problematic for L2 measures aggregated from L1 informant responses. The use of group means obtained by aggregating sampled L1 informant responses to represent L2 constructs assumes no measurement or sampling error, which can introduce bias in estimates of between-level indirect effects. This is more pragmatic when non-randomly sampling a small number of individuals within each L2 units or working with not highly reliable scales. From a multilevel perspective, observed L1 ratings should not be directly used to represent either the L2 construct or individual differences in perceptions of climate (Morin et al., 2014) because failure to disentangle sources of the trait and residual variances can introduce bias in estimates of indirect effects at either level.

Second, MLM approaches conflate the within-group(L1) and between-group(L2) effects through the use of an average slope estimate that can lead to biased indirect effects because the conflated slope estimation may contain effects that are unrelated to the mediation effects (Preacher et al., 2010). For example, in a 2-1-1 mediation design, where predictor (X) is measured at L2 and both the mediator (M) and outcome (Y) are measured at L1, the indirect effect only exists at L2 because variation in X is a within-cluster constant and cannot differentially influence individual differences within a cluster on the mediating variable (Hofmann, 2002). In other words, while the effect of X on M, and the effect of X on Y, are unbiased in a traditional MLM frameworks, the effect of M on Y (i.e., the 1-1 link) conditional on X should contain two separate elements: one strictly at L2 and one strictly at L1. When L1 and L2 effects of an L1 predictor on an L1 outcome are not equal, the use of this conflated slope to compute the indirect effect (a*b) will produce bias in the L2 indirect effects (Preacher et al., 2010; Raudenbush & Bryk, 2002). More generally, in any mediation model involving at least one variable measured at L2, the indirect effect only occurs at L2, and the use of combined L2 and L1-cluster slopes in any design involving a 1-1 linkage (e.g., X-1-1, 1-1-Y) will lead to bias in estimates of indirect effects (Preacher et al., 2010). In fact, in multilevel contexts, use of a single mean slope will misrepresent the data and result in a biased inference at either level (Preacher et al., 2010).

An unconflated multilevel modeling (UMM) approach (e.g., Zhang, Zyphur, & Preacher, 2009) has been proposed to separate L1 and L2 effects by replacing the single predictor $X_{ij}$ with a cluster mean centered ($X_{ij}$ -$X_{.j}$) predictor and the group mean $X_{.j}$, where the former represents the L1 proportion of $X_{ij}$ and the latter represents the L2 proportion of $X_{ij}$ (Preacher et al., 2010). However, concern remains with this approach as it makes use of the observed group mean as a proxy for a cluster measure in MLM, which tends to bias the L2 effects (Preacher et al., 2010). This bias is particularly pronounced with small numbers of items and is not improved even with large sample sizes (Muthén, Muthén, & Asparouhov, 2016). Research has shown that the UMM approach still returns potentially biased estimates of between indirect effects when using observed group means as proxies for true group measures in any mediation design with a 1-1 linkage (e.g., X-1-1, 1-1-Y; Lüdtke et al., 2008; Preacher et al., 2010; Preacher, Zhang & Zyphur, 2011).

Third, MLM approaches are limited in applications with upper-level mediators or outcomes (e.g., 1-2-1, 2-2-1, 1-1-2; Preacher, 2015; Preacher et al., 2010), as they require that the dependent variable in each causal link be assessed at L1 (e.g., 1-1-1, 2-1-1; Krull & Mackinnon, 2001). Although several alternative approaches for modeling upper-level dependent variables have been described in the literature (e.g., two-step analyses, aggregation and disaggregation; see Preacher et al., 2010), they bring their own unique limitations.

**Doubly Latent Multilevel Mediation Models in a MSEM Framework**

In addressing the limitations of using observed variable MLM methods to model mediation hypotheses with nested data, Preacher et al. (2010) presented a general multilevel structural equation model (MSEM) that integrates CFA/SEM and MLM into a single framework. The main advantage of this approach is that it decomposes the observed L1 ratings into two unrelated L1 and L2 latent variance components, thereby separating L1 and L2 effects and removing bias in estimates of the L2 indirect effect that is inherent in the use of the observed cluster means as L2 measures in traditional MLM approaches (Lüdtke et al., 2008; Marsh et al., 2009, 2012; Preacher et al., 2010).

The MSEM framework decomposes L1 ratings into two orthogonal latent parts through extension of single level classical test theory, where an observed score is the sum of a true score ($\eta_x$) and measurement error ($R_x$), to multiple level contexts. Where, following Marsh et al. (2009), the equation for each indicator can be specified as:

$$X_{kij} = \mu_x + \lambda_{kb} * \eta_{xj} + \lambda_{kw} * \eta_{xij} + R_{kxj} + R_{kxij} \qquad \text{k=1} \sim \text{K items.} \qquad \text{[Equation 3]}$$

Where, K is the number of items, $\mu_x$ is the grand mean of observed variable X; $\lambda_{kb}$ and $\lambda_{kw}$ are the L2 and L1 factor loadings, respectively; $\eta_{xij}$ and $\eta_{xj}$ are the true scores at L1 and L2 that capture true L1 and L2 variation, respectively (Tofighi & Thoemmes, 2013); and $R_{xij}$ and $R_{xj}$ are the error scores at L1 and L2, respectively.

It is worth noting that the error score at L2 ($R_{xj}$) includes two components: 1) unreliability in the indicators used to assess their corresponding L2 construct, measurement error, and 2) unreliability of the observed cluster mean due to the sampling of a small number of individuals from each L2 unit (Lüdtke et al., 2011), sampling error. The MSEM framework expands the single-level structural equation model (SEM) into multilevel measurement and structural components by adding random coefficients to the SEM equations. This allows a subset of parameters to randomly covary across clusters, and the modeling of covariances in the L2 portion of the structural model. The simplified MSEM model can be written as three matrix equations (Preacher et al., 2010):

| | | |
|---|---|---|
| Measurement model | $\mathbf{Y}_{ij} = \mathbf{\Lambda}\boldsymbol{\eta}_{ij}$ | [Equation 4] |
| Structural model L1 | $\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_{\mathrm{j}} + \mathbf{B}_{j}\boldsymbol{\eta}_{ij} + \boldsymbol{\zeta}_{ij}$ . | [Equation 5] |
| Structural model L2 | $\boldsymbol{\eta}_{j} = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_{\mathrm{j}} + \boldsymbol{\zeta}_{\mathrm{j}}$ . | [Equation 6] |

Where, $\mathbf{Y}_{ij}$ is a $p$ dimensional vector of observed variables across K items; $\mathbf{\Lambda}$ is a $p \times m$ loading matrix, where $m$ is the number of latent factors decomposed across L1 and L2; and $\boldsymbol{\eta}_{ij}$ is a $m$ vector of latent factors.

In the L1 structural model, $\alpha_j$ and $\mathrm{B}_j$ contain intercepts and regression slopes linking latent components, respectively. Elements of the vector $\alpha_{\mathrm{j}}$ ($m \times 1$) and matrix $\mathrm{B}_j$ ($m \times \mathrm{m}$) can potentially vary at the cluster level, so the vector $\eta_j$ ($r \times 1$) contains random effects for all elements in Equation 5, where $r$ is the number of random effects. Notably, $\eta_j$ is very different from $\eta_{ij}$, in that $\eta_{ij}$ consists of all the decomposed latent components. The vector μ ($r \times 1$) and matrix β ($r \times r$) contain fixed effects. Specifically, μ consists of means of random effects and intercepts of L2 structural equations. β contains regression slopes among random effects. In a mediation design, β would contain the disaggregated L2 effects. $\zeta_{ij}$ and $\zeta_{\mathrm{j}}$ are residuals at L1 and L2, respectively.

As a result, the multilevel mediation equations in a general MSEM framework can be succinctly specified as follows when all variables are measured at L1:

Measurement model: k = 1 ~ K items. [Equation Set 7]

$$Y_{kij} = \eta_{kYj} + \eta_{kYij} = \mu_Y + \lambda_{kbY} * \eta_{Yj} + \lambda_{kwY} * \eta_{Yij} + R_{kYj} + R_{kYij}$$
$$M_{kij} = \eta_{kMj} + \eta_{kMij} = \mu_M + \lambda_{kbM} * \eta_{Mj} + \lambda_{kwM} * \eta_{Mij} + R_{kMj} + R_{kMij}$$
$$X_{kij} = \eta_{kXj} + \eta_{kXij} = \mu_X + \lambda_{kbX} * \eta_{Xj} + \lambda_{kwX} * \eta_{Xij} + R_{kXj} + R_{kXij}$$

L1 structural model: [Equation set 8]

$$\eta_{Yij} = c'_w \eta_{Xij} + b_w \eta_{Mij} + \zeta_{Yij}$$
$$\eta_{Mij} = a_w \eta_{Xij} + \zeta_{Mij}$$
$$\eta_{Xij} = \zeta_{Xij}$$
$$\eta_{Yj} = \alpha_{\eta Yj}$$
$$\eta_{Mj} = \alpha_{\eta Mj}$$
$$\eta_{Xj} = \alpha_{\eta Xj}$$

L2 structural model: [Equation set 9]

$$a_w = \mu_{aw} + \zeta_{aw}$$
$$b_w = \mu_{bw} + \zeta_{bw}$$
$$c'_w = \mu_{c'w} + \zeta_{c'w}$$
$$\alpha_{\eta Yj} = \mu_{Yj} + c'_b \eta_{Xj} + b_b \eta_{Mj} + \zeta_{Yj}$$
$$\alpha_{\eta Mj} = \mu_{Mj} + a_b \eta_{Xj} + \zeta_{Yj}$$
$$\alpha_{\eta Xj} = \mu_{Xj} + \zeta_{Xj}$$

The L2 indirect effect can be obtained as the product of $a_b$ and $b_b$, and the L1 indirect effect as the product of $a_w$ and $b_w$. However, the L1 indirect effect computation would be different depending upon whether the L1 slopes are fixed or random (Preacher et al., 2010). Comparing the MSEM equations (i.e., Equations 7-9) with the MLM equations (i.e., equations 1-2), the main advantages of MSEM mediation analysis include: 1) decomposing the L1 variable into two latent components so that potential indirect effects at each level can be estimated (Preacher et al., 2010, 2011); 2) adding the random intercepts as latent means to represent the between component of the L1 variable to avoid biasing the L2 indirect effect caused through the use of observed cluster means (Preacher et al., 2011; Preacher, 2015); and 3) extending the single-level SEM model to a MSEM adds an additional structural equation at the cluster level that contains all random intercepts and slopes. This third advantages allows cluster level regressions among the random coefficients on each other so that the model can accommodate L2 variables as outcomes in the causal chain. Finally, MSEM can provide fit indices to evaluate absolute and relative model fit (Preacher et al., 2010).

The MSEM approach can be used to define a doubly latent construct (Marsh et al., 2009, 2012; Lüdtke et al., 2011) based on multiple indicators---latently aggregating multiple items into latent factors at both levels as in traditional CFA models, and latently aggregating L1 ratings within the same cluster to form latent L2 aggregates (i.e., individual responses in the same cluster serve as multiple indicators of a latent L2 construct). The latent aggregation process from L1 to L2 occurs by modeling the intercepts of the regressions of observed indicators on latent factors at L1 as random intercepts (i.e., varying across clusters; Marsh et al., 2009), thus correcting for the sampling error arising from the selection of L1 units within L2 clusters. Here, the L2 construct is doubly latent in the sense of simultaneously correcting for measurement error due to the sampling of items and persons while aggregating L1 into L2 (Marsh et al., 2012; Morin et al., 2014).

**Substantive Illustration of Doubly Latent Multilevel Mediation Models in a MSEM Framework**

We illustrate the usefulness of this approach in investigating the mediating role of engagement in associations between reports of school climate and a variety of student outcomes in schools. Four models were examined in turn to consider the four outcomes of the Prevalence of Teasing and Bullying (PTB), student GPA, short-term suspension rates, and long-term suspension rates. A large body of research has demonstrated the importance of a positive school climate as being critical to the prevention of peer aggression and victimization experiences (e.g., Bradshaw, Waasdorp, & Johnson, 2015; Cook, Williams, Guerra, Kim, & Sadek, 2010; Cornell, Shukla, & Konold, 2015; Gregory, Fan, Sheras, Shih, & Huang, 2010; Johnson, 2009; Swearer, Espelage, Vaillancourt, & Hymel, 2010), promotion of academic achievement (e.g., Konold, Cornell, Jia, & Malone, 2018; Lee, 2012; Wang & Degol, 2016), and reduction of disciplinary consequences (e.g., Gregory, Cornell, & Fan, 2011). Establishing a functional model of how school climate influences student outcomes has both theoretical and practical implications (Konold et al., 2018). Research has also begun to explore the possibility that school climate might indirectly impact student and school outcomes through a variety of intermediate mechanisms (Benbenishty, Astor, Roziner, & Wrabel, 2016) that could hold the key to developing more effective interventions for fostering school improvement efforts (Loukas, 2007). Student engagement is one of the candidate variables that could link school climate and student outcomes. For example, studies have shown that better student engagement leads to student academic success and less problematic behaviors (e.g., Lawson & Masyn, 2015; Wang & Eccles, 2013; Wang & Fredricks, 2014). Other studies have found that students more engage in schools characterized by more positive climate (e.g., Berkowitz, Moore, Astor, & Benbenishty, 2017; Cornell, Shukla, & Konold, 2016; Lee, 2012; Wang & Eccles, 2013). Konold and his colleagues (2018) have found that authoritative schools (i.e., high structure and high support) were associated with better student academic achievement, and student engagement serving as one of the mediating mechanisms. Another study also found student engagement mediates the positive relationship between classroom climate and academic achievement (Reyes, Brackett, Rivers, White, & Salovey, 2012). Informed by these studies, we hypothesize that student engagement might be one of the mechanisms that can explain the relationship between school climate and student academic, behavioral and disciplinary outcomes. That is, positive school climate advance better student academic, behavioral and disciplinary outcomes via promoting student engagement.

School climate is often said to capture the overall quality and character of school life (Cohen et al., 2009). As a result, it is appropriate that measurements of this construct include the perspectives of those that spend large amounts of time in their schools. Although judgments of school climate are typically obtained on the basis of individual perceptions of these informants, school climate can be considered a school-level construct (Konold & Cornell, 2015) that reflects the collective experiences of its constituents (Wang et al., 2014). In the present study, we obtained measures of school climate (i.e., fairness, justness, and support) through teacher ratings, and measures of engagement and student outcomes of teasing and bullying (PTB) and self-report GPA in schools through student ratings. Different informant types were used to help alleviate potential concerns with shared method variance (Wang & Degol, 2016) and common informant-based effects (Konold & Cornell, 2015). Since school climate is conceptualized as a school-level construct, primary interest was in the L2 indirect effects of engagement on relationships between school climate and student outcomes. L2 school demographic covariates are also included in the doubly latent multilevel mediation illustration, see Figure 1, taking PTB as an example.

The substantive research question of whether school climate factors are indirectly related to student outcomes through student engagement lends itself particularly well to illustrating the doubly latent multilevel mediation model for three reasons. First, there is a nested data structure as the school-level constructs were measured through individual reports obtained by informants within the schools. Second, each factor was measured with multiple observed variable indicators, and the student and teacher informants were sampled from within their schools. The doubly latent approach allows for the control of biases in parameter estimates that can arise from these sources of potential measurement error and sampling error, respectively. Third, design requirements for the doubly latent models typically require a relatively large number of L2 clusters (e.g., 100 clusters) and a sufficient number of nested L1 respondents (e.g., at least 10) in each L2 unit to optimize model convergence (Morin et al., 2014). Our sample satisfies these recommendations with 296 schools and an average cluster size of 201 students and 38 teachers.

# Method

## Data Sources

The sample drew from a statewide assessment of school climate and safety conditions administered to all 322 State public schools in spring 2016 (Cornell et al., 2016). Of the 322 schools eligible for participation, teacher surveys were received from 302 schools and student surveys were received from 320 schools. Schools could choose the random sampling option (i.e., invite 25 randomly selected students from each grade) or the whole grade option in which all students were invited to participate. Students completed the survey in classrooms under teacher supervision through a secure online Qualtrics platform. Teachers completed the survey voluntarily, and all teachers were invited to participate.

Data screening included identifying survey response times that were too fast, failing built-in validity check items (see Cornell et al., 2016 for additional details), removing schools with less than 3 teacher responses, and those with reports from only students or teachers. The resulting analytic sample consists of 59,581 students and 11,336 teachers from 296 schools. The student sample (51.3% female) was distributed across 9th (27.5%), 10th (25.9%), 11th (24.6%), and 12th (22.0%) grades. The race/ethnicity breakdown of students was 54.5% White, 17.6% Black, 11.8% Hispanic, 5.9% Asian, 1.0% American Indian or Alaska Native, 0.5% Native Hawaiian or Pacific Islander, with an additional 8.7% reporting more than one race. Teacher respondents were predominantly female (67.0%) and were 83.1% White, 8.3% Black, 3.4% Hispanic, 1.3% Asian, 0.2% American Indian or Alaska Native, 0.1% Native Hawaiian or

Pacific Islander, and 3.4% were classified as other/two or more races.

The 296 schools had an average enrollment of 1,397 students (range 59 to 4,190). Information from Virginia Department of Education records indicated that the average percentage of students eligible for free or reduced-price meals (FRPM) in the participating schools was 38.9%, with a range of 1.5% to 100.0%. The average percentage of White students across all the participating schools was 59.6%, with a range of 0.4% to 100.0%.

**Measures**

We focused on three key domains assessed on the Secondary School Climate Survey: School Climate, Student Engagement, and student self-reported outcomes, including the Prevalence of Teasing and Bullying (PTB) and self-reported GPA. Other objective school characteristics (%White, %FRPM, school size) and disciplinary outcomes were obtained from the State department of education records.

**School Climate-teacher reports**. The teacher version of the authoritative school climate survey included items measured on a 6-point agreement scale with response options ranging from *strongly agree* to *strongly disagree.* A comprehensive psychometric analysis (Huang & Cornell, 2016) of high schools revealed strong support for three school climate factors when evaluated through multilevel CFA: Fairness (5 items), Justness (3 items), and Support (10 items) with CFI = .92, TLI = .91, RMSEA = .05, $SRMR_{L1}$ = .04, and $SRMR_{L2}$ = .09. Moreover, school-level reliability estimates for these three factors were .92, .70, and .79, respectively.

**Student reports of engagement.** Student engagement in school was measured with six items derived from the Commitment to School Scale (Thornberry, Lizotte, Krohn, Farnworth, & Jang, 1991). The scale has two aspects of student engagement, affective and cognitive engagement. A*ffective engagement* refers to the student's positive feelings toward school, such as liking school and feeling belong to the school. C*ognitive engagement* concerns the student's investment in learning at school. The exemplar of the items are "I like this school", "I feel like I belong at this school", "I usually finish my homework", and "I want to learn as much as I can at school", see Konold et al., 2014). Each component was measured by three items with 4 response categories (*1= "Strongly disagree", 2= "Disagree", 3= "Agree", 4= "Strongly agree"*). Higher scores reflected greater levels of student engagement at school. A previous study with 39,364 middle school students (Konold et al., 2014) supported two factors at both levels in a multilevel confirmatory analysis context, with CFI = .99, TLI = .99, RMSEA = .05, $SRMR_{L1}$ = .03, and $SRMR_{L2}$ = .20. Reliability estimates for two factors at student level were .85 and .66, .87 and .96 at the school level, respectively in the middle school sample. In the present study, reliability estimates for affective and cognitive engagement are .89 and .70 at the student level, and .96 and .69 at the school level, respectively.

**Self-report academic achievement.** The survey asked "What grades did you make on your last report card?" The seven response options on this item ranged from "Mostly A's" to "Mostly D's and F's". Student responses were recoded so that students with "Mostly A's" scored at 7, "Mostly A's and B's" scored at 6, "Mostly B's" scored at 5, and so on, with a response of "Mostly D's and F's" scored at 1.

**Prevalence of Teasing and Bullying (PTB)--student reports.** A five-item scale was used to measure PTB on a 4-point agreement scale. Support for the internal and external validity of the scale has been demonstrated across multiple samples as well as multi-groups for gender and minority status (Bandyopadhyay et al., 2009; Klein, Cornell, & Konold, 2012; Mehta, Cornell, Fan, & Gregory, 2013). Multilevel CFA results with a statewide sample of students yielded good fit at both the student and school level (Konold et al., 2014). All factor loadings for student level (> .69) and school-level (> .81) were large and statistically significant, and reliability estimates for students and schools in a statewide sample of high schools were .85 and .93, respectively (Konold & Cornell, 2015). In the present study, observed scale scores ranged between 5 and 20, with reliability estimation of .86 at the student level and .96 at the school level.

**Disciplinary outcomes**. Student level suspension records from the state department of education for 2015-16 school year were deidentified and aggregated at school level to serve as external continuous school-level disciplinary outcomes. Within each school, unduplicated counts of students who have been referred to short-term suspension (STS) or long-term suspension (LTS) were weighted by their school size to obtain the short-term suspension and long-term suspension rate.

**Analytic Plan**

Grahic representations of the doubly latent multilevel models are illustrated in Figures 1-3. For ease of illustration,  the engagement measure in the model is shown as a single factor. However, all investigated models treated engagement as a two factors in order to separately evaluate its cognitive and affective components.  With substantive interest and focus at the school level, we describe salient features of the PTB outcome model, see Figure 1. PTB was regressed on the three school climate factors (paths c1 to c3) as well as Engagement (path b). In addition, Cognitive and Affective Engagement were regressed on the three school climate factors (paths a1 to a3), and the indirect effects (product of a and b) of the school climate factors on PTB through Engagement were measured to assess the potential processes underlying these linkages. Also shown in the L2 school portion of Figure 1 are three school demographic variables (i.e., school enrollment size, %FRPM, %White) that served as covariates on the PTB outcome. Given the focus of L2 school effects, no structure was imposed on the L1 factors and correlations among them were freely estimated.

In order to examine the amount of school-level PTB variance that could be explained by the school climate and engagement factors, over that which could be explained by the school covariates, a two-step model assessment approach was adopted. In step 1, paths a1-a3, c1-c3, and b were constrained to zero in order to assess the amount of variance in PTB that could be explained by the covariates. Thereafter in step 2, all paths were freely estimated. Partial mediation effects were tested by simultaneously estimating the direct paths from the school climate factors to PTB, as well as the indirect paths from the climate factors to PTB through engagement (Morin et al., 2014). In addition, based on theoretical and empirical evidence (Konold, et al., 2014), the residual covariances between two engagement aspects and covariances between three climate factors were specified. WLSMV was selected as the estimator because the majority of item indicators are categorical in nature (Beauducel & Herzberg, 2006; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Analyses were implemented through Mplus 7.3 in the context of multilevel structural equation modeling framework. All school-level constructs are doubly latent in the sense of simultaneously correcting for measurement error and sampling error due to the sampling of items and persons when aggregating L1 into L2 (Marsh et al., 2012; Morin et al., 2014).

Goodness of fit was evaluated with the Tucker-Lewis index (TLI), comparative fit index (CFI), root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) (Browne & Cudeck, 1993; Hu & Bentler, 1995). Values of .90 or greater (Hu & Bentler, 1999) on the CFI and TLI, and RMSEA and SRMR less than .08 or less than .06 are often taken as evidence of adequate or excellent fitting models, respectively.

Statistical tests of mediating effects have received increased attention in recent years due to the typically asymmetric sampling distributions of the product terms that are used to evaluate whether they are statistically greater than zero (Darlington & Hayes, 2017). One common approach to overcoming this problem involves the use of bias-corrected bootstrap confidence intervals (see for example, Lau & Cheng, 2012). However, this procedure is not currently available in Mplus for multilevel structural models. As a result, we used a Monte Carlo (MC) based parametric bootstrap approach that makes no assumptions about the sampling distribution of the product term (Preacher et al., 2010, 2011). R code for estimating 95% confidence intervals (CI) for the indirect effects was generated through the use of the Monte Carlo Method for Assessing Mediation (MCMAM; Selig & Preacher, 2008) software tool. Confidence intervals for the indirect effects were obtained on the basis of asymptotic variances and covariance's of the unstandardized parameter estimates of the direct effects surrounding a given mediating effect, and were based on 20,000 replications.

## Results

Descriptive statistics for school covariates and observed variable school-level factor means are presented in Table 1. Intraclass correlations for all item level indicators in the present study ranged from .02 to .20, and intraclass correlations for the school climate factors, engagement factors, and PTB factor ranged from .01 to .21. All L2 school reliabilities that reflect agreement among the L1 informants were appreciable for the three school climate factors of fairness (.91), justness (.65), and support (.75) that were estimated from teacher reports; and the affective engagement (.96), cognitive engagement (.69) and PTB (.94) factors measured by student reports.

Standardized model coefficients and fit statistics are shown in Table 2 for both the reduced (step 1) and full (step 2) doubly latent multilevel mediation models. Estimation of the model at step 1 where only the effects of the school covariates on outcomes were estimated resulted in poor fit for the L2 school portion of the model (e.g., with the outcome of PTB, CFI = .97, TLI = .97, RMSEA = .01, $SRMR_{L1}$ = .07, $SRMR_{L2}$ = .28). In this step 1 model, the percentage of White students was positively associated with PTB ($p < .01$), self-reported GPA ($p < .001$), while negatively associated with short-term suspension ($p < .05$); and the percentage of students receiving free and reduced-priced meals was significantly positively associated with student reports of PTB ($p < .001$), short-term suspension ($p < .001$), and long-term suspension ($p < .05$), but negatively associated with GPA ($p < .01$), together, they explained the L2 variance in student outcomes, ranging from 11% to 37%.

**PTB**

Model fit at step 2, where the direct and indirect effects were estimated, revealed substantial improvement over the step 1 model, with L2 school-level fit being adequate, for the outcome of PTB, CFI = .97, TLI = .97, RMSEA = .01, $SRMR_{L1}$ = .07, $SRMR_{L2}$ = .08. Higher levels of teacher perceived support in schools was associated with higher levels of student reported affective engagement ($B = .53$, $p < .001$) and cognitive engagement ($B = .45$, $p < .05$). Although justness was found to be directly associated with affective engagement ($B = .23$, $p < .05$) cognitive engagement was not. Higher levels of affective engagement in schools were associated with reports of less teasing and bullying ($B = - .64$, $p < .001$). No direct association was found between cognitive engagement and PTB by taking account of affective engagement. Teacher perceived support was both indirectly associated with PTB via affective engagement ($b = -0.27$, $p < .01$, 95%MC CI = -0.47, -0.09) and directly associated with PTB ($B = -.38$, $p < .01$). Although justness was not found to be directly related to PTB ($B = .05$, $p > .05$), it was indirectly associated with PTB through affective engagement (*b* = -0.20, $p < .05$, 95%MC CI = -0.39, -0.04), indicating that higher levels of teacher reported support or justness in schools were associated with students reporting more affective engagement, which in turn resulted in less teasing and bullying. In combination, the school climate factors and engagement accounted for an additional 56.9% of variance in PTB reports beyond the L2 school control variables.

**Short-term Suspension**

Model fit at step 2 with the outcome of short-term suspension (STS) was favorable, CFI = .96, TLI = .95, RMSEA = .02, $SRMR_{L1}$ = .08, $SRMR_{L2}$ = .08. Higher levels of teacher perceived support in schools was associated with higher levels of student reported affective engagement (*B* = .49, $p < .001$) and cognitive engagement (*B* = .43, $p < .001$). Justness was found to be directly associated with affective engagement (*B* = .23, $p < .05$) but not cognitive engagement. Higher levels of affective engagement in schools was associated with reports of lower short-term suspension rate (*B* = -.29, $p < .001$). However, no direct association was found between cognitive engagement and STS after controlling for affective engagement.

Indirect effects revealed that teacher perceived justness was both indirectly associated with STS via affective engagement (*b* = -0.02, $p < .05$, 95%MC CI = -0.04, -0.005) and directly associated with STS (*B* = -.26, $p < .01$). Although teacher perceived support was not found to be directly related to STS (*B* = .10, $p > .05$), it was indirectly associated with STS through affective engagement (*b* = -0.03, $p < .01$, 95%MC CI = -0.05, -0.01). The significant indirect effect indicated that higher levels of teacher reported support or justness in schools were associated with students reporting more affective engagement, which in turn resulted in lower short-term suspension rate. In combination, the school climate factors and engagement accounted for an additional 14.9% of variance in STS beyond the L2 school control variables.

**Long-term Suspension**

Model fit at step 2 with the outcome of long-term suspension (LTS) was favorable, CFI = .96, TLI = .95, RMSEA = .02, $SRMR_{L1}$ = .08, $SRMR_{L2}$ = .08. Higher levels of teacher perceived support in schools was associated with higher levels of student reported affective engagement ($B$ = .49, $p$ < .001) and cognitive engagement ($B$ = .43, $p$ < .001). Justness was found to be directly associated with affective engagement ($B$ = .23, $p$ < .05) but not cognitive engagement. Higher levels of cognitive engagement in schools was associated with lower long-term suspension rates ($B$ = -.27, $p$ < .05). No direct association was found between affective engagement and LTS after controlling for cognitive engagement.

Teacher perceived support was indirectly associated with LTS via cognitive engagement ($b$ = -0.003, $p$ < .05, 95%MC CI = -0.04, -0.005) but not directly associated with LTS ($B$ = -.21, $p$ > .05). No mediation effects were found for the other school climate factors. The significant indirect effect indicated that higher levels of teacher reported support in schools were associated with students reporting more cognitive engagement, which in turn resulted in lower long-term suspension rates. In combination, the school climate factors and engagement accounted for an additional 14.7% of variance in LTS beyond the L2 school control variables.

**Self-Reported GPA**

Model fit at step 2 with the outcome of self-reported GPA was favorable, CFI = .97, TLI = .96, RMSEA = .02, $SRMR_{L1}$ = .08, $SRMR_{L2}$ = .08. Higher levels of teacher perceived support in schools was associated with higher levels of student reported affective engagement ($B$ = .53, $p$ < .001) and cognitive engagement ($B$ = .44, $p$ < .05). Justness was found to be directly associated with affective engagement ($B$ = .23, $p$ < .05) but not cognitive engagement. Higher levels of cognitive engagement in schools was associated with higher self-reported GPA ($B$

= .84, $p < .001$). No direct association was found between affective engagement and GPA after controlling for cognitive engagement.

Indirect effects revealed that teacher perceived support was indirectly associated with self-reported GPA via cognitive engagement ($b = 0.29$, $p < .05$, 95%MC CI = 0.01, 0.69) but not directly associated with self-reported GPA ($B = -.25$, $p > .05$). No mediation effects were found for school climate factors of fairness and justness. The significant indirect effect indicated that higher levels of teacher-reported adult support in schools was associated with students reporting more cognitive engagement, which in turn resulted in higher self-reported GPA in schools. In combination, the school climate factors and engagement accounted for an additional 53.4% of variance in GPA beyond the L2 school control variables.

## Discussion

Doubly latent MSEM is an important and currently underutilized tool in applications of multilevel modeling in general, and when assessments of indirect effects are of interest. The approach is particularly useful in applications where L1 observations are sampled from within L2 units (e.g., a subsample of student or teacher informants within schools) to estimate cluster-level constructs. We illustrated the usefulness of this procedure in the context of several substantive examples related to school climate research. The model included the estimation of indirect effects and covariates on the student outcomes.

In contrast to traditional MLMs, it allowed for control of measurement error with respect to the sampling of items, and sampling error in relation to the sampling of informants within schools, to provide unbiased parameter estimates (Preacher et al., 2010). This occurs through decomposition of the observed variables at L1 into L1 and L2 latent components. As a result, potential indirect effects at each level can be estimated separately, thereby avoiding biased L2 indirect effects that can occur when aggregating observed variable L1 observations to form L2 cluster means in traditional multilevel models (Preacher, 2015). In observed variable MLMs, bias results from the implicit assumption that there is no measurement or sampling error present when aggregating sampled L1 responses to form L2 group means. From a multilevel perspective, observed L1 ratings include both L1 and L2 variances, so simply manifestly aggregating all the L1 responses to form higher-level constructs would not well represent the true measure of higher-level constructs. The L1 responses should not be directly used to represent either the L2 construct or individual differences in perceptions of the measured constructs (Morin et al., 2014) because failure to disentangle sources of the trait and residual variances can have an adverse impact on estimates of indirect effects at either level.

Three benefits of using the doubly latent model in applications of mediation analysis were illustrated with multiple outcomes measured at different levels and with different types of scales: (1) it takes account of both measurement and sampling error when aggregating L1 responses to form L2 constructs, reducing estimation bias for substantive associations. (2) It is able to estimate level specific effects without introducing bias by separating observed L1 variable into straight L1 and L2 variances. (3) The model can easily accommodate any outcomes and mediators assessed at either level.

**Reliability of L2 Latent Traits**

Reliability in the resulting L2 latent trait constructs regarding to multiple items can be assessed by way of traditional tools (e.g., coefficient alpha) and estimates of model fit. In addition, the reliability of L2 trait factors with respect to sampling persons can be assessed by level of agreement among the L1 observations that form them and L2 cluster size:

$$\frac{\tau_x^2}{\tau_x^2 + (\delta_x^2/n_j)}$$

Where, $\tau_x^2$ = between cluster variance, $\delta_x^2$ = within cluster variance, and $n_j$ is the average size of the L2 units (Lüdtke et al., 2011; Morin et al., 2014).

This reliability estimate of the aggregated L2 construct, obtained from L1 ratings, varies as a function of the average agreement among informants in the same L2 unit which can be evaluated by the intraclass correlation coefficient (ICC) and the number of sampled informants in L2 units (Marsh et al., 2012). The higher ICC, the larger the number of informants in L2 units, and the better reliability of the L2 aggregates. In the current model, the reliabilities, in relation to sampling error, of the aggregated L2 school climate factors of justness (.65), support (.75), and fairness (.91) were generally lower than the L2 factors of student reported engagement (.96) and PTB (.94). One of the likely reasons for this is because the average cluster size for students (i.e., *n* = 201) was much larger than that for teachers (i.e., *n* = 38). At the same time, with the same average cluster size for teachers, the school-level aggregate of fairness had a larger reliability estimate than either support or fairness. This difference can be attributed to stronger agreement among teacher perceptions of fairness in the same school (*ICC* = .21).

School aggregates of cognitive engagement had relatively low sampling reliability due to a very small ICC. Thus, when small numbers of L1 informants are sampled from L2 units, and the agreement among the L1informants is low, the sampling reliability of the L2 factors obtained from L1 informants will be low. In these instances, the need for doubly latent models that control for sampling error, and other sources of measurement unreliability is particularly important.

On the other hand, the sampling reliabilities of the L2 student factors of affective engagement and PTB in the present example were close to 1, indicating the near absence of sampling error. For these factors, controlling for the sampling error in forming L2 factors from individual L1 informants was not as important. The need for doubly latent models may also diminish when the *n/N* sampling ratio approaches 1 (i.e., when all L1 members of L2 clusters are sampled). Here, some would argue that when this ratio is 1, it is reasonable to assume that there is no sampling error and that the L2 construct can be represented by manifestly aggregating L1 variables (Marsh et al., 2012). Others (e.g., Hutchison, 2007; Shin & Raudenbush, 2010), however, argue that if each cluster is considered to be a sample from a larger population of informants, it would always be appropriate to control for sampling error through doubly latent models. Consequently, the decision to control for sampling error depends upon the nature of the research question (e.g., generalizability of the results, focus of the study) and also the sampling ratio (Marsh et al., 2012).

**Interpreting Multilevel Mediation Results**

Multilevel models in general provide opportunities to examine relationships among variables that may differ at various levels of an organizational hierarchy (e.g., students, schools, districts). Because constructs may have different meanings at different levels of analysis (Muthén, 1991), it is important to choose variables and constructs that make sense at the levels of intended interpretation. For example, some constructs might only be meaningful on an individual level (e.g., personality) *or* a school level (e.g., racial diversity). By contrast, other constructs are likely to have meaning at a variety of levels. Examples of these might include the substantive variables examined in the current illustration, where the constructs are likely to hold importance for both students and the schools they attend. For example, student perceptions of engagement and PTB are likely to be influenced by the overall levels of these features within a given school, and both student and school measures of these dimensions may have different causes and influences.

One of the advantages of a doubly latent model over a MLM is that it avoids conflation of the L1 and L2 effects that can otherwise bias indirect effects. When the predictor, mediator, and outcome in a mediation design are all observed at L1, unbiased estimates of both L1 and L2 indirect effects can be obtained to address mediation questions that are indicated from a given theoretical framework (Preacher et al., 2010) and in which the mechanisms of mediation may or may not be expected to differ across levels (Hofmann & Gavin, 1998). However, when using L1 variables that result from informant-based observations, it remains important to consider the substantive meaning of the resultant measures in relation to the referent being used and whether the variation makes sense at both levels of a mediation model (Preacher et al., 2010; Marsh et al., 2012). For example, student engagement in the current study was measured with self-referential items (e.g., I like this school, I usually finish my homework). As a result, it can potentially capture variation at both L1 and L2 with its own meaning at both levels. However, many of the items used to obtain measures of school climate in the current study were more school-referential in nature (e.g., most teachers and other adults at this school care about all students, students know the school rules for student conduct). Here, in order to measure a collective group construct (i.e., school climate), the individual L1 responses are important to form the L2 construct, but the L1 variation after abstracting the agreement among individuals within the school may not be as substantively meaningful as L2 climate measures and its effects on other variables. As a result, we focused our illustration on L2 mediation effects at the school level. A second reason for our L2 focus in the current applications was because we were unable to match individual student and teacher reports to one another due to the anonymous nature of survey that was intended to preserve confidentiality.

Given this L2 focus, and lack of L1 structure, the L1 climate factors (i.e., residuals) were correlated with the L1 mediator and outcome to obtain a saturated L1 model (Marsh et al., 2012;

Morin et al., 2014). After modeling the L2 structure as shown in Figure 1, the L1 residual correlations between ratings of school climate (i.e., deviations between individual teacher ratings of their school and the average school climate rating of that school) with engagement (mediator) and PTB (outcome) were non-significant and near zero. These results provide further confirmation that treating school climate as L2 constructs was appropriate for the current research design, and is consistent with Marsh et al. (2012) where the L1 residual ratings of classroom climate had no substantive effect on student outcomes after appropriately modeling classroom climate as a group construct.

**Implications for School Improvement Intervention**

Results of our models across academic, behavioral, and disciplinary outcomes revealed that schools described by teachers as being more supportive had higher levels of both affective and cognitive engagement, as perceived by students. However, only higher levels of cognitive engagement  were associated with higher levels of student achievement and lower levels of long-term suspension rates. Schools characterized as more just in their discipline had higher levels of affective engagement among students, but not cognitive engagement, and higher levels of affective engagement were associated with lower levels of teasing and bullying and short-term suspension rates.

Moreover, mediation results indicated that schools characterized by supportive adults could potentially reduce the level of student teasing and bullying directly, or indirectly by way of changing students' affective engagement in schools. Similarly, schools characterized by disciplinary justness could potentially reduce short-term school suspension rates directly, or indirectly by way of promoting students' affective engagement in schools. On the other hand, the mediation results showed that schools with more adult support could indirectly promote student achievement and reduce long-term suspension by way of increasing student cognitive engagement in schools.

These findings have important implications for school improvement on student achievement, bullying prevention, and disciplinary outcomes intervention. For instance, results provide evidence to support the idea that improvements in school climate can advance student engagement and therefore promote better student outcomes and reduce problematic behaviors (Bandyopadhyay et al., 2009; Konold et al., 2018; Loukas, 2007). Results of our analyses also lend continued support to the idea that students' engagement in schools is an important predictor of a variety of student outcomes (Bandyopadhyay et al., 2009; Fredricks, Blumenfeld, & Paris, 2004; Konold & Shukla, 2017; Konold et al., 2018).

Notably, two aspects of student engagement did not function the same way in terms of different outcomes we investigated in the current study. In some cases, cognitive engagement seems to be more important than affective engagement in the mediating process, and in other cases this was reversed. Cognitive engagement seems to be more associated with student achievement and long-term suspension, while affective engagement seems to be more associated with short-term suspension and the prevalence of teasing and bullying. The findings were consistent to Fredricks et al. (2004)'s arguement that different types of engagement "are likely to have different antecedents and outcomes" (p. 61). One potential explanation of the stronger association between cognitive engagement with long-term suspension rather than short-term suspension is that students who receive long-term suspension would less likely to engage in their cognitive learning because of the longer abscence from school. The limitation of the cross-sectional data would not allow us to answer the causal relationship. Further investigation and theoretical explanations are needed to be able to generalize theses findings to other settings and grades.

In addition, different school climate components were found to function differently for different student outcomes. In the current study, no additional effect of the fairness of school climate on all the investigated outcomes and engagement was found after controlling for the school climate factors of justness and support. Adult support, however, was found to play an important role in promoting student engagement and thereby better student outcomes across those investigated here. These findings also emphasize the critical role of teacher-student relationship in the secondary school context in promoting better student outcomes (e.g., Berkowitz et al., 2017; Jia, Konold, & Cornell, 2016; Quin, 2017).

**Conclusion and Future Directions**

The illustrations of doubly latent models in multilevel mediation analysis, with several outcome variables (both single indicator and multiple indicators) assessed at both L1 and L2, highlight the adaptability of this approach. The Mplus syntax used in generating the results of our models is provided as an Appendix to facilitate use of this procedure by interested readers. We also acknowledge that applications of doubly latent models typically require a relatively large number of L2 clusters (e.g., 100 clusters) and a sufficient number of nested L1 respondents (e.g., at least 10) in each L2 unit to optimize model convergence (Morin et al., 2014). Similarly, although foundational work on the conceptual and theoretical aspects of doubly latent models has been described (Marsh et al., 2009, 2012; Preacher et al., 2010, 2011), and simulations on specific forms of doubly latent approaches have been conducted (Koch & Semmler-Busch, 2019), more simulation work in this area is needed to better understand the behavior of model estimates under different design conditions; more simulation work on comparison of doubly latent approach and manifest approach is needed in order to provide guidance for applied researchers. Applications of this doubly latent procedure into multilevel moderation analysis would be the promising future directions too. It is also important to recognize that no statistical modeling approach can in itself establish causal inferences, and that such claims related to mediational relationships may require longitudinal data (Maxwell & Cole, 2007). Finally, although teacher perceptions of school climate provide different insights from student perceptions of school climate functioning, future studies might apply this procedure to integrate both student and teacher/staff perceptions of school climate as a more comprehensive measure of school climate and study its effects on a variety of outcomes.

## References


Bandyopadhyay, S., Cornell, D. G., & Konold, T. R. (2009). Validity of three school climate scales to assess bullying, engagement, and help seeking. *School Psychology Review, 38*, 338-355.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11*, 142–163. https://doi.org/10.1037/1082-989X.11.2.142

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186–203. https://doi.org/10.1207/s15328007sem1302_2

Benbenishty, R., Astor, R. A., Roziner, I., & Wrabel, S. L. (2016). Testing the causal links between school climate, school violence, and school academic performance: A cross-lagged panel autoregressive model. *Educational Researcher, 45*, 197–206. https://doi.org/10.3102/0013189X16644603

Berkowitz, R., Moore, H., Astor, R. A., Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research, 87*, 425–469. https://doi.org/10.3102/0034654316669821

Bradshaw, C. P., Waasdorp, T. E., & Johnson, S. L. (2015). Overlapping verbal, relational, physical, and electronic forms of bullying in adolescence: Influence of school context. *Journal of Clinical Child & Adolescent Psychology, 44*, 494–508. https://doi.org/10.1080/15374416.2014.893516

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, *111*, 180–213. Retrieved from https://schoolclimate.org/climate/documents/policy/School-Climate-Paper-TC-Record.pdf

Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly, 25*, 65-83. http://dx.doi.org/10.1037/a0020149

Cornell, D., Huang, F., Datta, P., Malone, M., Jia, Y., Burnette, A.G., Shukla, K., Konold, T., & Meyer, P. (2016). *Technical Report of the Virginia Secondary School Climate Survey: 2016 Results for 9th–12th Grade Students and School Staff.* Charlottesville, VA: Curry School of Education, University of Virginia.

Cornell, D., & Huang, F. (2016). Authoritative school climate and high school student risk behavior: A cross-sectional multi-level analysis of student self-reports. *Journal of Youth and Adolescence*, *45*, 2246–2259. https://doi.org/10.1007/s10964-016-0424-3

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

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis.* Stanford, CA: Stanford University Evaluation Consortium.

Darlington, R.B., & Hayes, A.F. (2017). *Regression Analysis and Linear Models: Concepts Applications, and Implementation*. New York: The Guilford Press.

Datta, P., Cornell, D., & Huang, F. (2016). Engagement and prevalence of bullying bystander behavior in middle school. *Psychology in the Schools*, *53*(8), 804-816.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, & Evaluation, 14(20)*, 1-11.

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*, 149-167. https://doi.org/10.1016/j.leaqua.2004.09.009

Every Student Succeeds Act (ESSA) of 2015, Pub. L. No 114-95 § 114 Stat. 1177 (2015–2016). https://www.congress.gov/bill/114th-congress/senate-bill/1177/text.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research, 74*, 59-109. https://doi.org/10.3102/00346543074001059

Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for Black and White high school students. *American Educational Research Journal, 48*, 904–934. https://doi.org/10.3201/002831211398531

Hofmann, D. A. (2002). Issues in multilevel research: Theory development, measurement, and analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 247–274). Malden, MA: Blackwell.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*, 623–641.

https://doi.org/10.1177/014920639802400504

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. https://doi.org/10.1080/10705519909540118

Huang, F. L., & Cornell, D. G. (2016). Multilevel factor structure, concurrent validity, and test–retest reliability of the high school teacher version of the Authoritative School Climate Survey. *Journal of Psychoeducational Assessment*, *34*, 536–549. https://doi.org/10.1177/0734282915621439

Huang, F. L., Cornell, D. G., & Konold, T. R. (2015). Engagement in middle schools: A factor structure and criterion-related validity study. *Assessment*, *22*, 497-512. https://doi.org/10.1177/1073191114551016

Hutchison, D. (2007). When is a compositional effect not a compositional effect?. *Quality & quantity*, *41*, 219-232. https://doi.org/10.1007/s11135-007-9094-2

Jia, Y., Konold, T. R., & Cornell, D. (2016). Authoritative school climate and high school dropout rates. *School Psychology Quarterly*, *31*, 289–303. https://doi.org/10.1037/spq0000139

Johnson, S. L. (2009). Improving the school environment to reduce school violence: A review of the literature. *The Journal of School Health, 79*, 451–465. http://dx.doi.org/10.1111/j.1746-1561.2009.00435.x

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8*, 115–128. https://doi.org/10.1037/1082-989X.8.2.115

Klein, J., Cornell, D., & Konold, T. (2012). Relationships between bullying, school climate, and student risk behaviors. *School Psychology Quarterly*, *27*, 154–169. https://doi.org/10.1037/a0029350Koch, T., & Semmler-Busch, U. (2019). On the power and performance of a doubly latent residual approach to explain latent specific factors in multilevel-bifactor-(S-1) models. Structural Equation Modeling: A Multidisciplinary Approach, 26, 212-235. https://doi.org/10.1080/10705511.2018.1516506

Konold, T. R., 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. https://doi.org/10.1037/spq0000062

Konold, T. R., & Cornell, D. (2015). Measurement and structural relations of an authoritative school climate model: A multi-level latent variable investigation. *Journal of School Psychology*, *53*, 447–461. https://doi.org/10.1016/j.jsp.2015.09.001

Konold, T., Cornell, D., Jia, Y., & Malone, M. (2018). School climate, student engagement, and academic achievement: A latent variable, multilevel multi-informant examination. *AERA Open*, 4(4). https://doi.org/10.1177/2332858418815661

Konold, T. R., & Shukla, K. (2017). Estimating school climate traits across multiple informants: An illustration of a multitrait–multimethod validation through latent variable modeling. *Educational Assessment*, *22*, 54–69. https://doi.org/10.1080/10627197.2016.1271705

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249-277. https://doi.org/10.1207/S15327906MBR3602_06

Lau, R.S., & Cheung, G.W. (2012). Estimating and comparing specific mediation effects in complex latent variable models. *Organizational Research Methods, 15*, 3-16. https://doi.org/10.1177/1094428110391673

Lawson M. A., Masyn K. E. (2015). Analyzing profiles, predictors, and consequences of student engagement dispositions. *Journal of School Psychology, 53*, 63–86. https://doi.org/10.1016/j.jsp.2014.11.004

Lee, J.S. (2012). The effects of the teacher–student relationship and academic press on student engagement and academic performance. International Journal of Educational Research, *53*, 330–340. http://dx.doi.org/10.1016/j.ijer.2012.04.006

Loukas, A. (2007). What is school climate. *Leadership Compass*, *5*(1), 1–3. Retrieved from http://laspdg.org/files/What%20is%20School%20Climate.pdf

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*, 444–467. https://doi.org/10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229. https://doi.org/10.1037/a0012869

Luo, Z., Breslau, J., Gardiner, J. C., Chen, Q., & Breslau, N. (2014). Assessing interchangeability at cluster levels with multiple-informant data. *Statistics in Medicine, 33,* 361-375. https://doi.org/10.1002/sim.5948

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological methods, 12*, 23-44. https://doi.org/10.1037/1082-989X.12.1.23

Macmann, G. M., Barnett, D. W., & Lopez, E. J.  (1993). The child behavior checklist/4-18 and related materials: Reliability and validity of syndromal assessment. *School Psychology Review, 22,* 322-333.

Mehta, S., Cornell, D., Fan, X., & Gregory, A. (2013). Bullying climate and school engagement in ninth grade students.  *Journal of School Health, 83*, 45-52. https://doi.org/10.1111/j.1746-1561.2012.00746.x

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47,* 106–124. https://doi.org/10.1080/00461520.2012.670488

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44,* 764–802. https://doi.org/10.1080/00273170903333665

Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *The Journal of Experimental Education, 82,* 143–167. https://doi.org/10.1080/00220973.2013.769412Muthén, B.O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28,* 338-354. https://doi.org/10.1111/j.1745-3984.1991.tb00363.x

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and mediation analysis* *using Mplus*. Los Angeles, CA: Muthén & Muthén.

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: The Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology*, *31*, 629–656. https://doi.org/10.1080/01443410.2011.586416

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual review of psychology*, *66,825-852*. https://doi.org/10.1146/annurev-psych-010814-015258

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233. https://doi.org/10.1037/a0020141

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: the advantages of multilevel sem. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(2), 161–182. https://doi.org/10.1080/10705511.2011.557329

Quin, D. (2017). Longitudinal and contextual associations between teacher–student relationships and student engagement: A systematic review. *Review of Educational Research*, *87*, 345-387. https://doi.org/10.3102/0034654316669434

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*, 354-373.

http://dx.doi.org/10.1037/a0029315

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*, 700-712. http://dx.doi.org/10.1037/a0027268

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353.

Selig, J.P., & Preacher, K.J. (2008). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer software]. Available from: http://quantpsy.org/.

Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics*, *35*, 26–53. https://doi.org/10.3102/1076998609345252

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. http://dx.doi.org/10.3102/0013189X09357622

Thornberry, T. P., Lizotte, A. J., Krohn, M. D., Farnworth, M., & Jang, S. J. (1991). Testing interactional theory: An examination of reciprocal causal relationships among family, school, and delinquency. *Journal of Criminal Law & Criminology*, *82,* 3–35. Retrieved from http://scholarlycommons.law.northwes tern.edu/jclc/vol82/iss1/2

Tofighi, D., & Thoemmes, F. (2014). Single-level and multilevel mediation analysis. *The Journal of Early Adolescence*, *34*, 93-119. https://doi.org/10.1177/0272431613511331U.S. Department of Education (2017). Revised state template for the consolidated state plan: The Elementary and Secondary Education ACT of 1965, as amended by the Every Student Succeeds Act. Retrieved from: https://www.ed.gov/essa?src=rn

Van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods*, *12*, 368–392. https://doi.org/10.1177/1094428107309322

Wang, M. T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*, 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wang M., & Eccles J. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28,* 12–23. https://doi.org/10.1016/j.learninstruc.2013.04.002

Wang, M. T., & Fredricks, J. A. (2014). The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence. *Child development*, *85,* 722-737. https://doi.org/10.1111/cdev.12138

Wang, W., Vaillancourt, T., Brittain, H. L., McDougall, P., Krygsman, A., Smith, D., … Hymel, S. (2014). School climate, peer victimization, and academic achievement: Results from a multi-informant study. *School Psychology Quarterly*, *29,* 360–377. https://doi.org/10.1037/spq0000084

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14,* 301–322. http://dx.doi.org/10.1037/a0016972

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12,* 695-719. https://doi.org/10.1177/1094428108327450

Table 1

*Descriptive Statistics for Observed Variables and Scales*

| | Mean | *SD* | Min | Max |
|---|---|---|---|---|
| School Size | 1227.700 | 734.627 | 59.000 | 4190.000 |
| % White | 59.580 | 26.510 | .448 | 100.000 |
| % FRPM | 38.877 | 20.634 | 1.530 | 100.000 |
| $\text{Fairness}_{\text{teacher}}$ | 3.903 | 0.525 | 2.570 | 5.350 |
| $\text{Justness}_{\text{teacher}}$ | 5.117 | 0.225 | 3.920 | 5.760 |
| $\text{Support}_{\text{teacher}}$ | 4.785 | 0.235 | 3.910 | 5.600 |
| $\text{PTB}_{\text{student}}$ | 2.495 | 0.201 | 1.790 | 2.990 |
| Affective $\text{engagement}_{\text{student}}$ | 2.948 | 0.232 | 1.700 | 3.460 |
| Cognitive $\text{Engagement}_{\text{student}}$ | 3.314 | 0.082 | 3.050 | 3.520 |
| Self-report GPA | 5.348 | 0.412 | 4.010 | 6.620 |
| STS rate | 0.091 | 0.100 | 0.001 | 1.099 |
| LTS rate | 0.005 | 0.014 | 0.000 | 0.202 |

Note. PTB = Prevalence of teasing and bullying, STS = Short Term Suspension, LTS = Long Term Suspension.
School size was divided by 100 in analyses to facilitate model convergence.
Reported scale descriptive statistics are based on observed variable school-level factor means.

Table 2
*Doubly Latent Multilevel Mediation Standardized Regression Coefficients*

| | Outcomes | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PTB | | | Self-report GPA | | | LTS | | | STS | | |
| | $\beta_{Step1}$ | $\beta_{Step2}$ | | $\beta_{Step1}$ | $\beta_{Step2}$ | | $\beta_{Step1}$ | $\beta_{Step2}$ | | $\beta_{Step1}$ | $\beta_{Step2}$ | |
| School Size | 0.069 | 0.073 | | 0.113 | 0.113 | | -- | -- | | -- | -- | |
| % White | 0.225 | 0.222 | ** | 0.471 | 0.471 | *** | -0.157 | -0.157 | | -0.167 | -0.167 | * |
| % FRPM | 0.414 | 0.410 | *** | -0.245 | -0.245 | ** | 0.234 | 0.234 | * | 0.485 | 0.485 | *** |
| ***Direct effect*** | | | | | | | | | | | | |
| Fairness --> AENG | 0.000 | -0.103 | | 0.000 | -0.100 | | 0.000 | -0.113 | | 0.000 | -0.113 | |
| Justness --> AENG | 0.000 | 0.229 | * | 0.000 | 0.228 | * | 0.000 | 0.231 | * | 0.000 | 0.229 | * |
| Support --> AENG | 0.000 | 0.532 | *** | 0.000 | 0.532 | *** | 0.000 | 0.486 | *** | 0.000 | 0.488 | *** |
| Fairness --> CENG | 0.000 | -0.087 | | 0.000 | -0.082 | | 0.000 | -0.088 | | 0.000 | -0.088 | |
| Justness --> CENG | 0.000 | -0.109 | | 0.000 | -0.011 | | 0.000 | -0.103 | | 0.000 | -0.102 | |
| Support --> CENG | 0.000 | 0.453 | * | 0.000 | 0.439 | * | 0.000 | 0.428 | * | 0.000 | 0.428 | * |
| AENG --> Outcome | 0.000 | -0.636 | *** | 0.000 | -0.220 | | 0.000 | -0.151 | | 0.000 | -0.290 | *** |
| CENG --> Outcome | 0.000 | 0.173 | | 0.000 | 0.838 | *** | 0.000 | -0.267 | * | 0.000 | 0.038 | |
| Fairness --> Outcome | 0.000 | 0.100 | | 0.000 | 0.035 | | 0.000 | -0.007 | | 0.000 | 0.124 | |
| Justness --> Outcome | 0.000 | 0.048 | | 0.000 | 0.229 | | 0.000 | -0.211 | | 0.000 | -0.262 | ** |
| Support --> Outcome | 0.000 | -0.379 | ** | 0.000 | -0.248 | | 0.000 | 0.203 | | 0.000 | 0.103 | |
| ***Indirect effect#*** | | | | | | | | | | | | |
| Fairness IND[A] | -- | 0.079 | | -- | 0.012 | | -- | 0.000 | | -- | 0.005 | |
| Justness IND[A] | -- | -0.200 | * | -- | -0.067 | | -- | -0.002 | | -- | -0.023 | * |
| Support IND[A] | -- | -0.265 | ** | -- | -0.091 | | -- | -0.002 | | -- | -0.029 | ** |
| Fairness IND[C] | -- | -0.008 | | -- | -0.037 | | -- | 0.000 | | -- | 0.000 | |
| Justness IND[C] | -- | -0.026 | | -- | -0.125 | | -- | 0.001 | | -- | -0.001 | |
| Support IND[C] | -- | 0.061 | | -- | 0.287 | * | -- | -0.003 | * | -- | 0.003 | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Chi-square* | 6821 | 6941 | | 7834 | 8215 | | 11286 | 12409 | | 10976 | 12163 | |
| *DF* | 828 | 817 | | 611 | 600 | | 568 | 557 | | 568 | 557 | |
| *RMSEA* | 0.011 | 0.011 | | 0.014 | 0.015 | | 0.018 | 0.019 | | 0.018 | 0.019 | |
| *CFI* | 0.969 | 0.968 | | 0.968 | 0.966 | | 0.960 | 0.955 | | 0.960 | 0.955 | |
| *TLI* | 0.966 | 0.965 | | 0.964 | 0.962 | | 0.956 | 0.950 | | 0.956 | 0.950 | |
| *SRMRw* | 0.065 | 0.065 | | 0.075 | 0.075 | | 0.077 | 0.077 | | 0.077 | 0.077 | |
| *SRMRb* | 0.312 | 0.081 | | 0.223 | 0.084 | | 0.202 | 0.080 | | 0.203 | 0.081 | |
| | | | | | | | | | | | | |
| School-level $R^2$ | 0.118 | 0.687 | *** | 0.366 | 0.900 | | 0.107 | 0.254 | *** | 0.325 | 0.474 | *** |
| $\Delta R^2$ | -- | 0.569 | *** | -- | 0.534 | *** | -- | 0.147 | *** | -- | 0.149 | *** |

Note. School Size = Enrollment / 100.
PTB = Prevalence of Teasing and Bullying, FRPM = Free or Reduced-Price Meals, LTS = Long-Term Suspension, STS = Short-Term Suspension, AENG = Affective Engagement, CENG = Cognitive Engagement.

* $p < .05$. ** $p < .01$. *** $p < .001$.
# Unstandardized results.
$IND^A$ Indirect effect through Affective Engagement, $IND^C$ Indirect effect through Cognitive Engagement

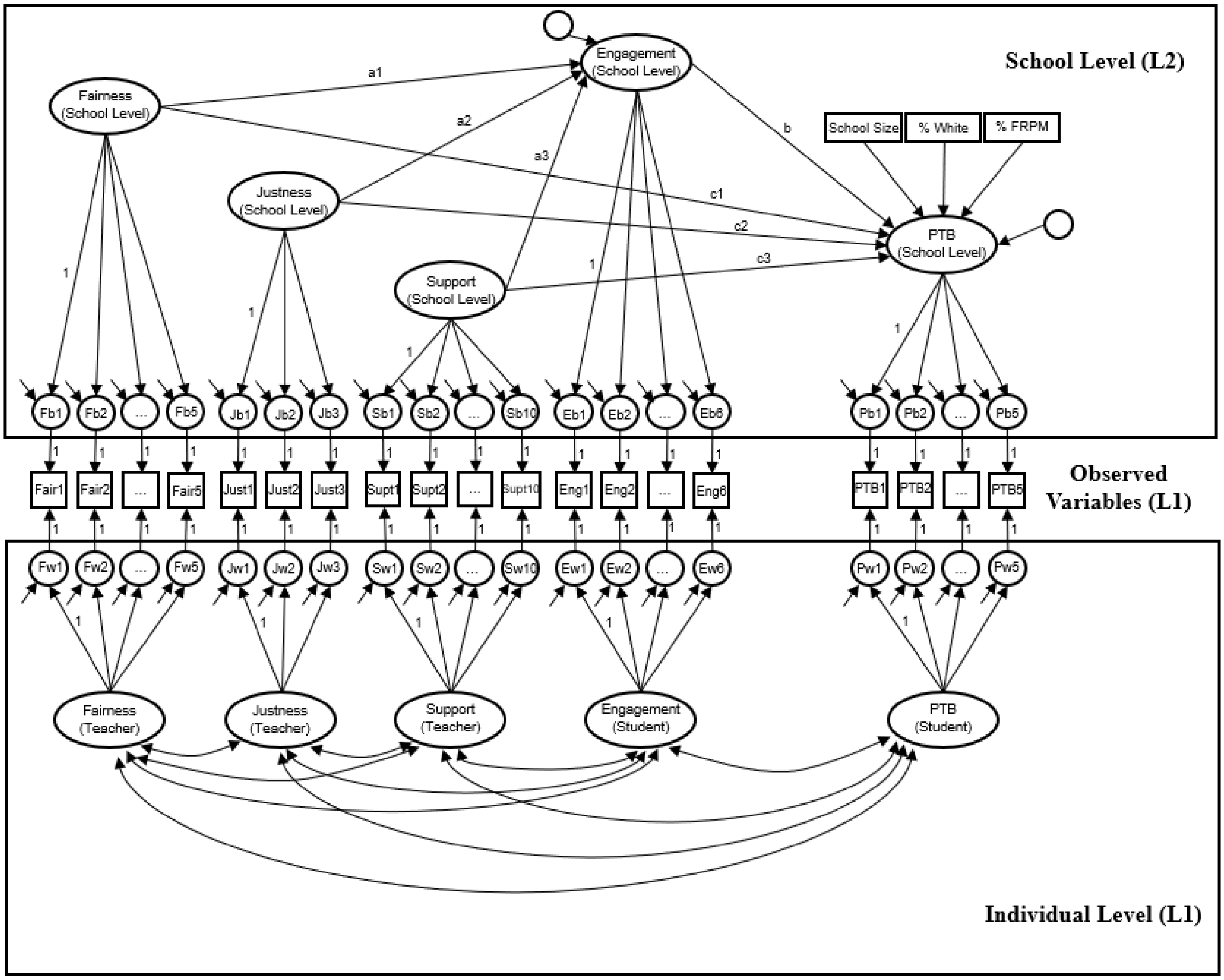


*Figure 1.* Doubly latent multilevel mediation model illustrating the mediating role of engagement in relationships among school climate factors and PTB

*Note.* Factor correlations among L2 school climate predictors of Fairness, Justness, and Support omitted for clarity of presentation.

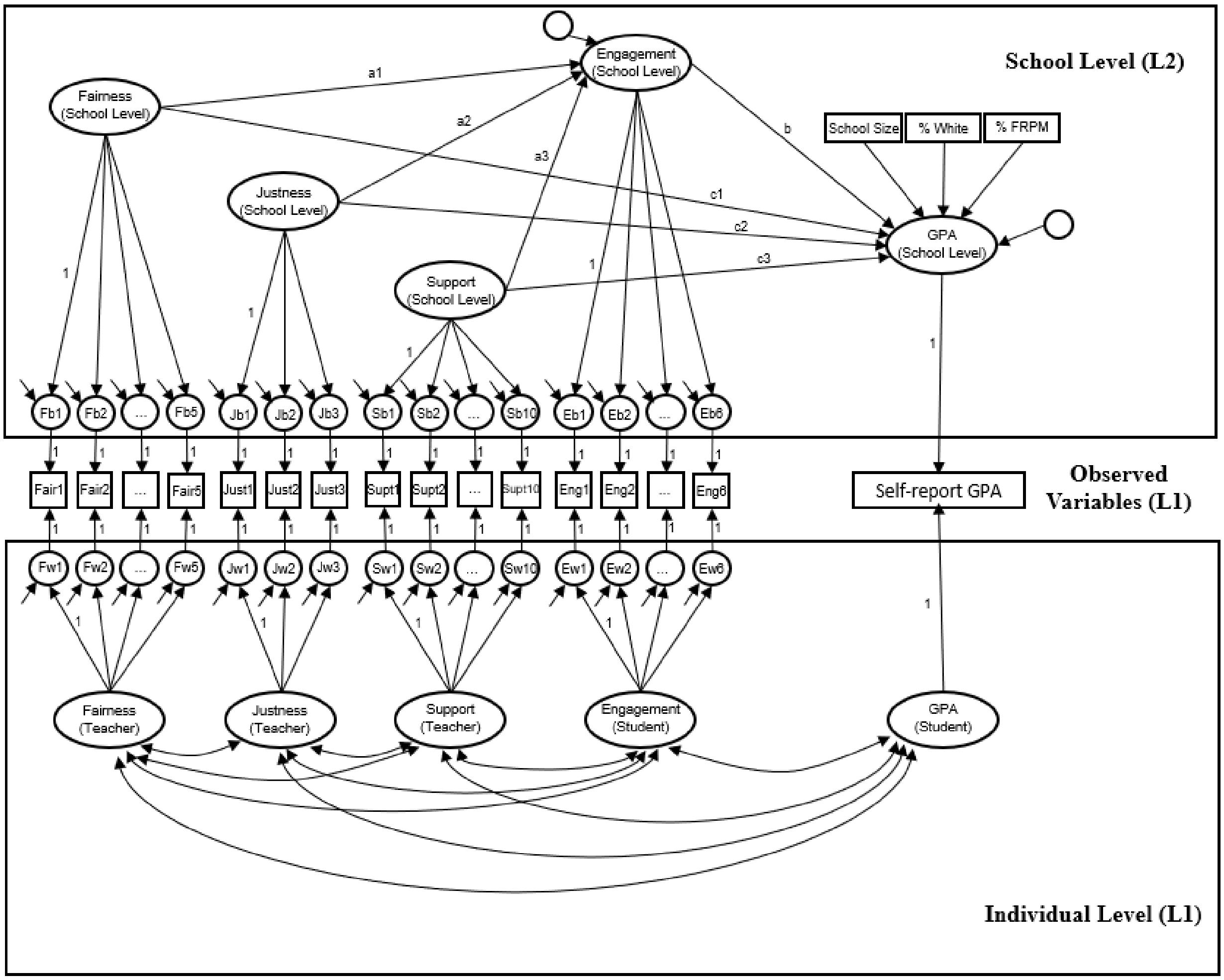


*Figure 2.* Doubly latent multilevel mediation model illustrating the mediating role of engagement in relationships among school climate factors and self-report GPA
*Note.* Factor correlations among L2 school climate predictors of Fairness, Justness, and Support omitted for clarity of presentation.

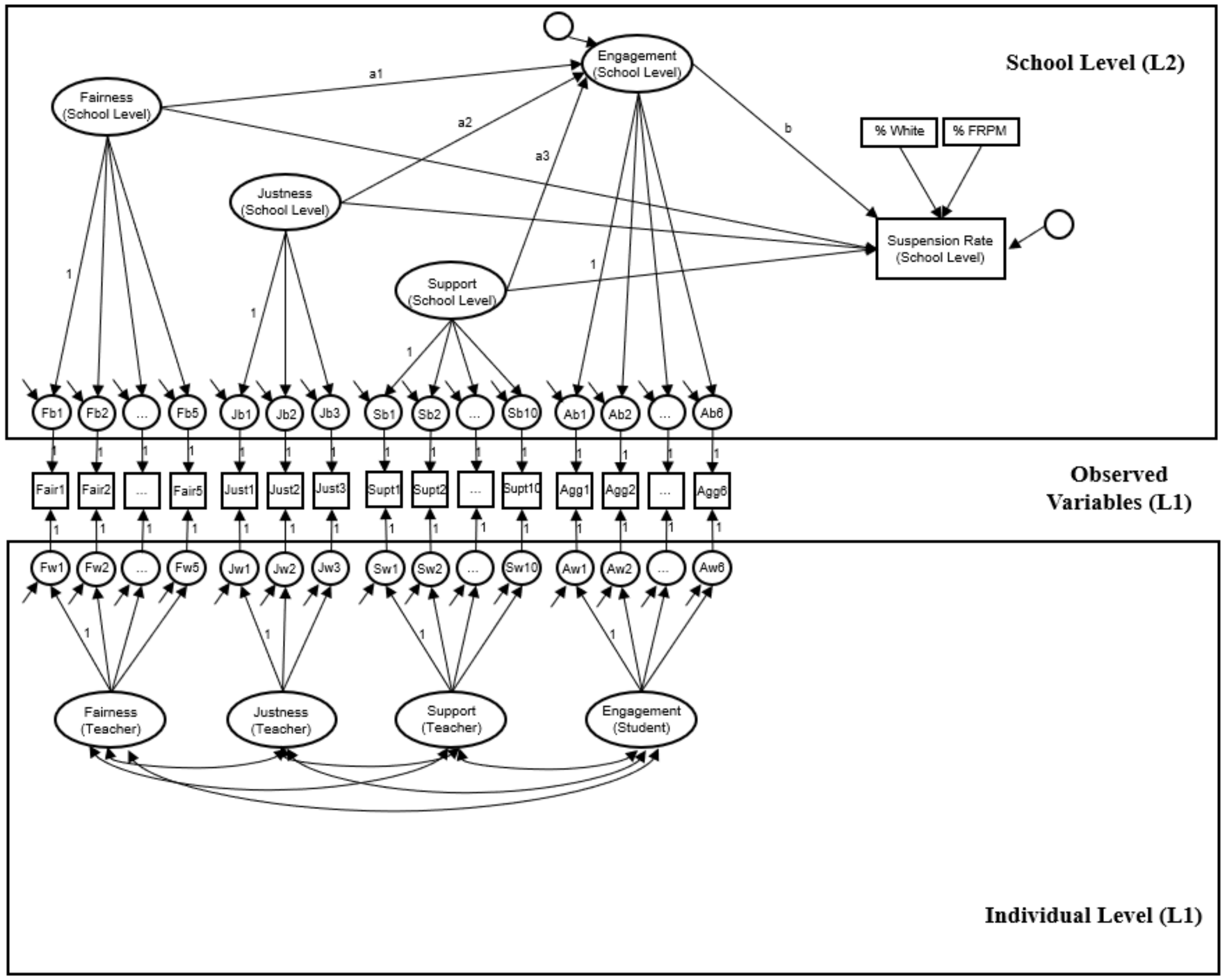


*Figure 3.* Doubly latent multilevel mediation model illustrating the mediating role of engagement in relationships among school climate factors and suspension rate

*Note.* Factor correlations among L2 school climate predictors of Fairness, Justness, and Support omitted for clarity of presentation.

**Appendix for Mplus syntax**:

TITLE: Doubly latent multilevel model of the mediating role of engagement in the association between school climate and PTB.

**DATA:** ! Statement to indicate the name and location of the data file
FILE IS mydata.csv;

**VARIABLE**: ! List, in order of appearance, the variables included in the data set.
**NAMES ARE** USID studid S4Div S4SchNum ENROL100 XX4RACWP XX4FRPMP
T4STR7 T4STR16 T4STR5 T4STR8 T4STR15R T4STR11R
T4STR13 T4STR10R T4SUP7 T4SUP8 T4SUP9 T4SUP10 T4SUP1 T4SUP2
T4SUP5 T4SUP6 T4SUP11 T4SUP3 S4PTB1 S4PTB2 S4PTB3 S4PTB9 S4PTB4
S4ENG1 S4ENG2 S4ENG8 S4ENG4 S4ENG7 S4ENG6;
**USEVARIABLES ARE** USID ENROL100 XX4RACWP XX4FRPMP
T4STR7 T4STR16 T4STR5 T4STR8 T4STR15R T4STR11R T4STR13
T4STR10R T4SUP7 T4SUP8 T4SUP9 T4SUP10 T4SUP1 T4SUP2 T4SUP5
T4SUP6 T4SUP11 T4SUP3 S4PTB1 S4PTB2 S4PTB3 S4PTB9 S4PTB4
S4ENG1 S4ENG2 S4ENG8 S4ENG4 S4ENG7 S4ENG6;
**CATEGORICAL** = S4PTB1 S4PTB2 S4PTB3 S4PTB9 S4PTB4 S4ENG1 S4ENG2 S4ENG8
S4ENG4 S4ENG7 S4ENG6 T4STR7 T4STR16 T4STR5 T4STR8 T4STR15R
T4STR11R T4STR13 T4STR10R T4SUP7 T4SUP8 T4SUP9 T4SUP10
T4SUP1 T4SUP2 T4SUP5 T4SUP6 T4SUP11 T4SUP3;

**CLUSTER** = USID; ! Specification of the clustering (school) variable

**BETWEEN**= ENROL100 XX4RACWP XX4FRPMP; ! L2 only variables

!Technical specifications
**ANALYSIS:**
TYPE = TWOLEVEL; ! Specifies a two level model
PROCESS = 4;
ESTIMATOR = wlsmv; ! Estimator specification

!Model specifications
**MODEL**:
**%Within%** ! Specification of the L1 portion of the model
! Teacher reports of Disciplinary Structure are based on two factors
T_FAIR_W BY T4STR7 T4STR16 T4STR5 T4STR8 T4STR15R;
T_JUST_W BY T4STR11R T4STR13 T4STR10R;

! Teacher reports of Support are based on one factor
T_SUPT_W BY T4SUP7 T4SUP8 T4SUP9 T4SUP10 T4SUP1 T4SUP2 T4SUP5
T4SUP6 T4SUP11 T4SUP3;

! Student reports of Engagement are based on one factor

```
    AENG_W BY S4ENG1 S4ENG2 S4ENG8;
    CENG_W BY S4ENG4 S4ENG7 S4ENG6;

   ! Student reports of PTB are based on one factor
   PTB_W BY S4PTB1 S4PTB2 S4PTB3 S4PTB9 S4PTB4;

  ! Specification of correlations among climate factors with mediator and outcome
   T_FAIR_W with AENG_W CENG_W PTB_W;
   T_JUST_W with AENG_W CENG_W PTB_W;
   T_SUPT_W with AENG_W CENG_W PTB_W;
   PTB_W with AENG_W CENG_W;
   AENG_W with CENG_W;


%Between% ! Specification of the L2 portion of the model
  ! Teacher reports of Disciplinary Structure are based on two factors (Predictors)
    T_FAIR_B BY T4STR7 T4STR16 T4STR5 T4STR8 T4STR15R;
    T_JUST_B BY T4STR11R T4STR13 T4STR10R;

  ! Teacher reports of Support are based on one factor (Predictor)
    T_SUPT_B BY T4SUP7 T4SUP8 T4SUP9 T4SUP10 T4SUP1 T4SUP2 T4SUP5
                T4SUP6 T4SUP11 T4SUP3;

  ! Student reports of Engagement are based on one factor (Mediators)
   AENG_B BY S4ENG1 S4ENG2 S4ENG8;
   CENG_B BY S4ENG4 S4ENG7 S4ENG6;
   AENG_B with CENG_B; ! Correlate the residuals of two mediators


  ! Student reports of PTB are based on one factor (Outcome)
   PTB_B BY S4PTB1 S4PTB2 S4PTB3 S4PTB9 S4PTB4;

   ! Regression paths
    PTB_B ON AENG_B(b1);
    AENG_B ON T_FAIR_B(a1);
    AENG_B ON T_JUST_B(a2);
    AENG_B ON T_SUPT_B(a3);

    PTB_B ON CENG_B(b2);
    CENG_B ON T_FAIR_B(a4);
    CENG_B ON T_JUST_B(a5);
    CENG_B ON T_SUPT_B(a6);


    PTB_B ON T_FAIR_B(c1);
    PTB_B ON T_JUST_B(c2);
```

```
    PTB_B ON T_SUPT_B(c3);

  ! Regression of PTB outcome on L2 covariates
    PTB_B ON ENROL100 XX4RACWP XX4FRPMP;

  ! Constraint on negative residual variances
   T4SUP9@0;
   S4ENG2@0;
   S4ENG8@0;
   T4STR8@0;
   T4STR16@0;
   T4STR13@0;
   T4STR11R@0;
   S4PTB9@0;
   T4STR7@0;

MODEL CONSTRAINT: ! Specification of indirect effects
   ! Indirect effects via affective engagement
    NEW (indb1);
    NEW (indb2);
    NEW (indb3);
    indb1=a1*b1;
    indb2=a2*b1;
    indb3=a3*b1;

  ! Indirect effects via cognitive engagement
    NEW (indb4);
    NEW (indb5);
    NEW (indb6);
    indb4=a4*b2;
    indb5=a5*b2;
    indb6=a6*b2;

OUTPUT: SAMPSTAT STDYX TECH1 TECH3 CINTERVAL;
```