Summer 8-11-2020

School Climate Methodology: Issues of Multilevel Data and Measurement Invariance

Faith A. Zabek
Georgia State University

ACCEPTANCE

This dissertation, SCHOOL CLIMATE METHODOLOGY: ISSUES OF MULTILEVEL DATA AND MEASUREMENT INVARIANCE, by FAITH ZABEK, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education & Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

Joel Meyers, Ph.D.
Committee Chair

Kenneth Rice, Ph.D.
Committee Member

Ann Kale Kruger, Ph.D.
Committee Member

Jeffrey Ashby, Ph.D.
Committee Member

Date

Brian Dew, Ph.D.
Chairperson, Department of Counseling and Psychological Services

Paul A. Alberto, Ph.D.
Dean, College of Education & Human Development

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education & Human Development's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

____

Faith Zabek

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Faith Ashley Zabek
Department of Counseling and Psychological Services
College of Education & Human Development
Georgia State University

The director of this dissertation is:

Joel Meyers
Department of Counseling and Psychological Services
College of Education & Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Faith Zabek

ADDRESS: 3841 Haulani Pl.
Princeville, HI 96722

EDUCATION:

Ph.D.	2020	Georgia State University Counseling and Psychological Services
M.Ed.	2014	Georgia State University School Psychology
B.A.	2010	DePauw University Psychology

PROFESSIONAL EXPERIENCE:

2018-Present	School Psychologist Hawai‘i Department of Education, Kaua‘i, HI
2019-Present	President Hawai‘i Association of School Psychologists
2017-2018	Doctoral Psychology Intern Hawai‘i Psychology Internship Consortium (HI-PIC: APA Accredited) Hawai‘i Department of Education, Kaua‘i, HI
2016-2017	Graduate Teaching Assistant, Counseling and Psychological Services Georgia State University, Atlanta, GA
2012-2017	Graduate Research Assistant, College of Education and Human Development Georgia State University, Atlanta, GA
2015	Psychological Assistant Georgia Pediatric Psychology, Atlanta, GA
2014-2015	School Psychology Advanced Practicum Student City Schools of Decatur, Decatur, GA
2013-2014	School Psychology Practicum Student Clayton County Public Schools, Riverdale, GA
2011-2012	Applied Behavioral Analysis Therapist Reinforcement Unlimited, Woodstock, GA
2010-2011	English Teacher Gumi Elementary School, Bundang-gu, South Korea

PRESENTATIONS AND PUBLICATIONS:

Wood, L., Roach, A. T., Kearney, M. A., & Zabek, F. (2018). Enhancing executive function skills in preschoolers through a mindfulness-based intervention: A randomized, controlled pilot study. *Psychology in the Schools, 55*(6), 644–660.

Tolleson, A., Grad, R., Zabek, F., & Zeligman, M. (2017). Teaching helping skills courses: Creative activities to reduce anxiety. *Journal of Creativity in Mental Health.* http://dx.doi.org/10.1080/15401383.2017.1281186

Kruger, A. C., Harper, E., Zabek, F., Collins, S., Perkins, C., & Meyers, J. (2016). Facilitating a school-based prevention of commercial sexual exploitation of children. *Health promotion practice, 17*(4). 530–536. http://dx.doi.org/10.1177/1524839916628863

Kruger, A. C., Zabek, F., Collins, S., Harper, E. A., Hamilton, C., McGee, M. C., ... & Meyers, J. (2016). African American girls' descriptions of life in high-risk neighborhoods. *School Psychology Forum, 10* (1), 29–40.

La Salle, T. P., Zabek, F., & Meyers, J. (2016). Elementary student perceptions of school climate and associations with individual and school factors. *School Psychology Forum, 10* (1), 55–65.

Zabek, F., Harper, E., Perkins, C., Meyers, J., & Kruger, A. C. (2017, February). *School-based prevention of CSEC and other gendered aggression*. Workshop presented to the National Association of School Psychologists Annual Convention, San Antonio, Texas.

Wood, L., Zabek, F., & Roach, A. (2017, February). *Improving children's executive functioning using mindfulness-based interventions*. Workshop presented to the National Association of School Psychologists Annual Convention, San Antonio, Texas.

Roach, A., Meyers, B., Perkins, C., Wood, L., Zabek, F., Kearney, M, Bialo, J. (2016, February). *Using Children's Literature to Introduce Mindfulness in the Classroom*. Workshop presented to the National Association of School Psychologists Annual Convention, New Orleans, Louisiana.

Zabek, F., Pedley, T., & Barbarasch, B. (2015, February). *Using your degree for diplomacy: Putting advocacy into action*. Workshop presented to the National Association of School Psychologists Annual Convention, Orlando, FL.

Zabek, F., Meyers, J., & La Salle, T. (2016, February). *A thankful climate: School climate, gratitude, and the influence of measurement invariance.* Paper presented to the National Association of School Psychologists Annual Convention, New Orleans, Louisiana.

Wood, L., Zabek, F., & Roach, A. (2016, February). *Mini-Mind: Mindfulness-based curriculum for preschool classrooms*. Paper presented to the National Association of School Psychologists Annual Convention, New Orleans, Louisiana.

La Salle, T., Zabek, F., & Meyers, J. (2014, February). *Examining relationships between school climate, culture, and ecology: Revisions to the Georgia Student Health Survey.* Paper presented to the National Association of School Psychologists Annual Convention, Washington, D.C.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2012-Present	*National Association of School Psychologists* (NASP)
2018-Present	*Hawaiʻi Association of School Psychologists* (HASP)
2017-2018	*Hawaiʻi Psychological Association* (HPA)
2015-2016	*American Educational Research Association* (AERA)

SCHOOL CLIMATE METHODOLOGY: ISSUES OF MULTILEVEL DATA AND MEASUREMENT INVARIANCE

by

FAITH ZABEK

Under the Direction of Joel Meyers

ABSTRACT

School climate has been recognized as an opportunity to foster student success due to its demonstrated links to desirable academic, social/emotional, and behavioral outcomes and its critical role in the school improvement process. The significance of school climate and the value of its study have been made clear both in educational literature and educational policy. As reflected in its inclusion in the Every Student Succeeds Act (ESSA) of 2015, more and more states are reporting school climate indicators alongside more traditional academic outcomes within their accountability systems. Accordingly, the stakes attached to the accurate measurement of school climate are greater than ever. Unfortunately, the complexity of school climate presents an array of challenges when attempting to measure it accurately. It is typically measured using survey data, from which several analytic issues arise. In particular, the clustered nature of survey data confounds the effects of school climate at individual and school levels. It is important that researchers clearly define the level of school climate being investigated and use appropriate statistical techniques to measure it. In addition, survey items and constructs may have different meanings for various groups of individuals within schools and across schools with differing characteristics – leading to invalid comparisons. Researchers should investigate the equality of school climate surveys for diverse student and school populations. This dissertation systematically reviews the techniques school climate researchers employ to address these issues during scale development. Then, it employs a bioecological framework to investigate the clustered nature and invariance of a school climate survey using multilevel confirmatory factor analysis and multilevel structural equation modeling procedures.

INDEX WORDS: School climate, multilevel methods, measurement invariance, systematic review, survey development, survey validation, multilevel CFA, multilevel MIMIC modeling

SCHOOL CLIMATE METHODOLOGY: ISSUES OF MULTILEVEL DATA AND MEASUREMENT INVARIANCE

by

FAITH ZABEK

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

School Psychology

in

the Department of Counseling & Psychological Services

in

the College of Education & Human Development
Georgia State University

Atlanta, GA
2020

DEDICATION

To my parents, Bob and Sandy,

my brothers, Josh and Drew,

my sisters, Bri and Tasha,

my nieces, Emerson, Kensington, Quinn, and Dylan,

and my other half, Kendall,

my gratitude can never be measured.

This is for you.

ACKNOWLEDGMENTS

This dissertation would not be possible without the guidance and support of Dr. Joel Meyers. I struck gold having you as my advisor. You motivated me to finish, and you inspire me to lead with humility, grace, and wisdom. Dr. Ken Rice provided invaluable assistance when conducting and interpreting the complex analyses presented in this dissertation. For me, the time that you both spent reviewing and revising multiple iterations of this document will forever be priceless. I would like to thank my other committee members, Dr. Ann Cale Kruger and Dr. Jeff Ashby, for your feedback and encouragement. I am in awe of your dedication to bettering the lives of children, youth, and the graduate students with whom you work. I would also like to thank James Pooley. You did not hesitate to take on the tedious task of coding the articles included in this review. I could not have done it without you. To Dr. Catherine Perkins, Dr. Kruger, Dr Meyers, and the rest of the Project P.R.E.V.E.N.T. team, my experience conducting culture-specific action research with you has kept me grounded in the lived experiences of students and the ultimate goals of research throughout this dissertation process. I am profoundly grateful. Finally, to the entire school psychology faculty at Georgia State University, your commitment to maintaining an exceptional program and preparing skilled school psychologists is incomparable. You made this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 MEASURES OF SCHOOL CLIMATE: A SYSTEMATIC REVIEW OF SURVEY VALIDATION METHODS

The importance of school climate was first introduced over a century ago. In 1882, philosopher Felix Adler emphasized the role of the *school atmosphere* throughout his book, *The Moral Instruction of Children*. He underscored that, to have a beneficial effect, the school atmosphere – comprised of the patterns of teaching, discipline, interactions, and relationships within schools – "should possess a sunny climate" (p. 33). Over 25 years later, a New York City school principal, Arthur C. Perry, expanded upon Adler's conception. Perry (1908) outlined how to cultivate a "sunny climate" in his publication, *The Management of a City School* (p. 261), where he suggested that principals must create the desired school atmosphere by encouraging a cheerful physical environment, quality teaching methods, and characteristics such as "fairness and justness" as well as "order, system, and neatness" (p. 262-263). Further elevating the function of the school atmosphere, psychologist and educational reformer John Dewey (1916) posited that education does not take place through merely the direct conveyance of knowledge. Instead, it occurs through the intermediary of the *school environment* – the conditions and social atmosphere that facilitate development, including: physical materials, methods of instruction, cooperative activities, social spirit, and shared experiences, attitudes, and values. He described "the school as a special environment," in which "the intermingling… of youth of different races, differing religions, and unlike customs creates for all a new and broader environment" (p. 25-26). The school environment, Dewey stressed, must be deliberately designed to coordinate the experiences of diverse groups for the purpose of education and growth.

These foundational ideas set the stage for the scientific study of school climate, which emerged with the rise of organizational climate research in the mid 20th century. In 1963, Halpin and Croft developed the Organizational Climate Descriptive Questionnaire to investigate the effects of school organizational climate on student learning and development (Halpin & Croft, 1963). Since then, numerous scholars have attempted to measure school climate and study its impact. Within the vast body of school climate research, unanimity regarding its specific definition has yet to be reached. However, the original interpretations introduced by Adler, Perry, and Dewey – that recognize the multidimensionality of school climate and its role in education – still radiate in modern conceptualizations. For example, the National School Climate Council (2007) describes it as the quality and character of school life, stating, "it is based on patterns of school life experiences and reflects norms, goals, values, interpersonal relationships, teaching, learning, and leadership practices, and organizational structures" (p. 5). Despite inconsistencies regarding its exact delineation, the significance of school climate relative to its association to student academic, behavioral, and social/emotional outcomes as well as to the school improvement process has been made clear not only in educational research (see, e.g., Thapa et al., 2013; Wang & Degol, 2016) but also in educational policy (see, e.g., the Every Student Succeeds Act [ESSA], 2015).

As the recognition of school climate's potential to enhance student success increases, so do efforts by researchers and educators to try to measure it. Wang and Degol (2016) found that roughly a quarter of the nearly 300 empirical school climate studies they reviewed dealt solely with the development and validation of surveys. Similarly, more and more states are attempting to evaluate school climate through statewide surveys and are reporting school climate indicators (e.g., the School Climate Performance Indicator: California Department of Education [CDE], 2019; the School Climate Star Rating index: Georgia Department of Education [GaDOE], 2015; the Strive HI School Climate measure: Hawai'i Department of Education [HIDOE], 2018) alongside their academic accountability systems (Education Commission of the States, 2018). Thus, the accuracy of school climate measurement is more important than ever. Unfortunately, the complexity of school climate presents an array of conceptual and statistical challenges when developing and choosing measures. For example, conceptual challenges arise due to inconsistent specifications of school climate dimensions (e.g. academic, community, safety), characteristics (e.g. affective, organizational), and theoretical frameworks. Statistically, challenges arise due to analytic issues related to the clustered nature of survey data and the need for equality of school climate measurement across diverse student and school populations (Bear et al., 2016; Konold et al., 2014; Konold & Cornell, 2015b; Wang & Degol, 2016; Zabek et al., 2017). Several reviews of school climate (Thapa et al., 2013; Wang & Degol, 2016) and its measurement (Kohl et al., 2013; Ramelow et al., 2015; Zullig et al., 2010) that have been published recently in refereed journals have directly addressed the conceptual issues described above related to the dimensions and characteristics of school climate, as well as the theoretical frameworks. However, there is a need for a systematic review of school climate measures with specific attention to the important statistical challenges related to clustered data and the need for equality of measurement across different groups.

Conceptual Issues

The conceptual issues described above lead to challenges when interpreting the meaning of school climate and its impact, as well as when making direct comparison across results obtained from different measures. A plethora of school climate measures have been used in research and practice (e.g., see, Clifford et al., 2012), and these instruments sometimes assess very different constructs. For example, two measures of school climate may include items that target different dimensions (e.g., social versus instructional dimensions) and different characteristics of interest (e.g., affective characteristics at the individual level versus organizational characteristics at the school level). Comparing results from these measures, without careful consideration of these conceptual inconsistencies, may lead to invalid conclusions. Due to the multifaceted nature of school climate, its measurement should be firmly grounded in theory (Ramelow et al., 2016). However, this is often not the case (Konold et al., 2014; Wang & Degol, 2016). The lack of theory-grounded measurement development in school climate research hinders the process that leads to scientific advancement (Goldhaber, 2000). A clear link between theory and methods of measurement is needed to formulate and test hypotheses, to further understanding of the relationships among school climate dimensions and their impact, and to create successful strategies for school improvement.

Recent reviews of school climate have organized their results according to these conceptual issues – highlighting patterns of previous research and offering suggestions for how to address these issues in the future. For example, several literature reviews have synthesized the conceptualizations of school climate in previous research and come to a general consensus regarding its essential domains: (a) Safety, (b) Academic/Teaching & Learning, (c) Community/Relationships, and (d) Institutional Environment (Cohen et al., 2009; Thapa, et al, 2013; Wang & Degol, 2016; Zullig et al., 2010). In addition, Wang & Degol (2016) thoroughly described different theoretical frameworks (e.g., bio-ecological: Bronfenbrenner & Morris, 2006; risk and resiliency: Zimmerman & Arunkumar, 1994; and stage-environment fit: Eccles et al., 1996) that support the inclusion of these dimensions and facilitate interpretation of school climate findings. Lastly, the specification of school climate as an affective, relational, and organizational construct has been discussed at length, and there is agreement that, while it is typically described as an organizational construct, it can be conceptualized at each of these levels (e.g., see Kohl et al., 2013; Wang & Degol, 2016). These reviews assist in integrating school climate research and provide a guide regarding best practices for those interested in studying it. While scholars should strive to be consistent in their conceptualization of school climate, it is of greater importance that researchers firmly ground their specifications in theory and that they clearly state these conceptualizations so that results can be interpreted appropriately.

Expanding upon the broad reviews of school climate literature, several recent reviews that focused specifically on school climate surveys have used the aforementioned conceptual issues to organize and evaluate measures. For example, Zullig et al. (2010) used the domains of school climate to organize measures. They selected the five most commonly cited student school climate surveys and evaluated each according to its appropriateness in measuring each dimension. Kohl et al. (2013) used Bronfenbrenner's (1977) ecological model to organize school climate measures according to the system each assessed (e.g. affective characteristics, organizational characteristics, and so on). They searched peer-reviewed journals and selected studies that used validated student school climate surveys to statistically analyze the link between school climate and aggression. Then, Kohl and colleagues provided short summaries of each survey according to the characteristics it assessed (e.g. interpersonal feelings, organizational patterns, and so on). Most recently, Ramelow et al. (2015) searched scientific journals and selected studies that, (1) measured at least two of the dimensions of school climate suggested by Cohen et al. (2009), and (2) tested the psychometric properties of school climate surveys used with middle and high school students. They then evaluated each survey in terms of its reliability and validity, the number of domains it captured, and the soundness of theoretical grounding in its development.

These previous reviews of student school climate surveys play an important role in advancing research and practice. They not only synthesize previous research practices and offer suggestions for improvement, but also serve as resources for scholars and educators seeking school climate measures that are appropriate for their needs. In particular, Ramelow et al.'s (2015) review provides a comprehensive evaluation of the quality of student school climate surveys in terms of both basic psychometric properties and conceptual issues like theoretical grounding and appropriate coverage of domains. However, each of these reviews focuses on surveys measuring student perceptions of school climate. School climate shapes the experiences and interactions of all school stakeholders, and surveys measuring the perceptions of adult stakeholders such as parents, teachers, and staff have yet to be integrated into peer-reviewed evaluations of school climate measures. In addition, these reviews do not capture trends in survey validation strategies that address the clustered nature of school climate data or that ensure the equality of measurement across diverse populations. To thoroughly summarize and evaluate practices in the development of school climate measures, there is a need for a systematic review that includes surveys measuring the perceptions of various stakeholders and that attends to advanced statistical issues associated with the data analysis.

Statistical Issues

After firmly grounding one's conceptualization of school climate in theory, it is critical to employ survey validation strategies that are appropriately linked to that conceptualization and that adequately address related statistical concerns. Beyond the general considerations regarding reliability and validity in measurement development (e.g. conventional internal consistency estimates and factor analyses), school climate survey development also necessitates consideration of the multilevel nature of the data (e.g. students clustered within schools) and the equality of the measurement model across diverse student and school populations (e.g. whether the survey measures school climate equally well for students from various racial/ethnic backgrounds). In contrast to the consideration paid to conceptual issues, reviews of school climate have afforded significantly less attention to these statistical challenges. In particular, the frequency with which researchers use validation strategies that address such challenges, and the types of techniques employed, have yet to be synthesized in reviews of school climate surveys.

Issues of Clustered Data

As described in the previous section, school climate can be characterized at various "levels" (e.g. affective, relational, organizational). For example, it can be conceptualized as an affective construct capturing personal perceptions at the individual level, or as an organizational construct capturing shared experiences at the school level. Researchers may be interested in investigating school climate at either, or both, of these levels. However, due to the clustered nature of individuals within schools, school climate survey data inherently contains components of both levels. For example, a student's responses to survey items are simultaneously affected both by personal factors and by the shared characteristics of their school. Therefore, responses on school climate surveys fail to meet the assumption of data independence and cannot be considered purely individual- or school-level variables (Bliese, 2000; Konold et al., 2014; Marsh et al., 2009, 2012; Morin et al., 2014; Muthén, 1991, 1994; Muthén & Asparouhov, 2011). To address these challenges, it is important that researchers not only conceptually identify the desired level(s) of analysis, but also statistically control for measurement and sampling error when conducting analyses. Historically, school climate researchers have often conducted analyses on survey data without controlling for such error (Schweig, 2014).

Several issues arise when researchers fail to appropriately account for clustered data. First, the construct of school climate may be very different at the individual and school level. Conceptually, it may have different meanings and may be comprised of different dimensions at each level. Statistically, the number of factors in a school climate survey and its overall factor structure, internal consistency estimates, and relationships with other constructs may vary between levels. Survey responses conflate the individual-level school climate phenomenon with the school-level phenomenon, which makes it difficult to interpret the meaning of results and hinders the identification of distinct findings at each level (Konold et al., 2014). Conflated findings may have adverse policy implications. For example, different dimensions of school climate may be defined at the individual and school levels. By using the incorrect level of analysis to determine dimensions, qualities of school climate may be incorrectly targeted in policy and practice, or may be missed altogether (Schweig, 2014). In addition, when survey responses are aggregated to the school level, sampling error associated with the varying number of participants within each school is rarely controlled for. Multilevel factor analytic and structural equation modeling techniques can be used to account for the clustered nature of survey data by disaggregating the individual- and school-level components (e.g., see, Marsh et al., 2009, 2012; P. Mehta & Neale, 2005; Muthén, 1991, 1994; Muthén & Asparouhov, 2011). While reviews of school climate literature have stressed the need for analyses that account for the hierarchical nature of survey data (e.g., see Ramelow et al., 2015; Wang & Degol, 2016), a systematic review of the use of such techniques within school climate measurement development is needed.

Issues of Invariance

Issues related to the equality of school climate measurement across diverse populations have arguably received even less attention in reviews of school climate literature than issues related to clustered data. School climate surveys are used to measure perceptions of numerous groups of people. Raters are males and females of various ages and racial/ethnic backgrounds within schools that have different economic resources and demographic compositions. It is important to ensure that surveys measure school climate equally well across these diverse groups, particularly when group comparisons are made. That is, it is important to ensure that the meaning of school climate – including the interpretations of and relationships among its dimensions and the items used to measure it – does not systematically and significantly vary based on group membership. For example, for two students with identical perceptions of school climate, the probability of an observed response to a survey item should be equal regardless of individual (e.g. gender or race/ethnicity) or school (e.g. economic resources or demographic makeup) factors (Kim et al., 2012).

In the past, school climate research often reported group similarities and differences in perceptions of school climate without first ensuring the equality of the construct and its measurement. Such findings may be the result of measurement error. Thus, comparisons and conclusions about group similarities or differences may not be valid. Statistically, measurement invariance procedures can be employed to determine whether the same construct is being measured, in the same way (i.e., without bias), for different groups (e.g., see, Chen, 2008; Jak, 2013). Encouragingly, testing invariance in relation to individual-level factors has become increasingly common in recent school climate practices (see, e.g., Bear et al., 2011, 2015; Konold et al., 2014; Zabek et al., 2016). However, we have found no examples of testing invariance of school climate measures in relation to school-level factors. In addition, the traditional measurement invariance procedures employed often fail to account for the clustered nature of the data. Invariance with regard to individual- and school-level groups should to be demonstrated using multilevel modeling techniques to ensure findings are psychometrically valid (see, e.g., Kim et al., 2012, 2015). A systematic review of school climate surveys with consideration of the use of invariance procedures would assist in revealing trends and identifying areas of needed improvement.

Present Review

High-quality measurement instruments for assessing school climate are a prerequisite for identifying associations between school climate and various outcomes of interest and for ensuring school climate findings are valid. The aim of this systematic review is to examine the conceptual and statistical trends within school climate measurement development. It will investigate the validation strategies used to develop surveys measuring various school stakeholders' perceptions of school climate, paying particular attention to the use of strategies that, (1) account for the clustered nature of school climate survey data, and (2) determine the invariance of surveys across diverse student and school populations. Failure to address these issues has increasingly been noted as weaknesses of school climate research (Dunn et al., 2015: Konold et al., 2017; Phillips & Rowler, 2015; Schweig, 2014) but has yet to be investigated in detail in a systematic review of school climate measures (Kohl et al., 2013; Ramelow et al., 2015; Zullig et al., 2010). In addition, current gaps in school climate measurement and best practices in instrumentation development and evaluation will be discussed.

Method

This systematic review of school climate survey validation strategies was performed according to the guidelines outlined within the Preferred Reporting Items for Systematic

Reviews and Meta-Analyses (PRISMA) Statement (Liberati et al., 2009). The PRISMA statement includes a 27-item checklist (see Table 1.1) and 4-phase flowchart (see Figure 1) that include the dimensions that are essential for optimal reporting of systematic reviews. Through careful adherence to the PRSIMA statement, the present review strives to ensure the clarity and transparency of reporting and bolster its value to researchers, educators, and policy makers.

Search Strategies

A literature search was conducted using the following terms: (“school climate” or “school environment” or “school atmosphere” or “school culture” or “educational climate” or “educational environment” or “educational atmosphere” or “educational culture”) and (“assess*” or “instrument*” or “inventory” or “measure*” or “scale” or “survey” or “test” or “tool” or “validation”) (*designates allowance of alternative word endings within search results). The search was conducted on March 11, 2018. Due to the changing definition of school climate and relative consensus of its dimensions in recent literature reviews (Cohen et al., 2009; Loukas, 2007; Thapa et al., 2013; Wang & Degol, 2016), as well as advancement in the requirements for scientific rigor in measurement development (Kallestad, 2010), the search was limited to the years 2007 through 2017 using EBSCO to search within the following databases: Academic Search Complete, Education Source, ERIC, PsycARTICLES, Psychology and Behavioral Sciences Collection, and PsycINFO. Additionally, the reference lists of identified articles and of previous reviews of school climate literature (e.g., Kohl et al., 2013; Ramelow et al., 2015; Zullig et al., 2010) were scanned to identify articles, which were then cross-referenced with the articles identified in the search (Liberati et al., 2009).

Inclusion Criteria

Studies were included if they: 1) are published in an English-language peer-reviewed journal between 2007 and 2017; 2) report on quantitative studies; 3) report the original development or refinement of tools that have been used to measure elementary, middle, and/or high school climate, as perceived by student, teachers/staff, or parents; 4) include at least two items and measure at least two dimensions of school climate; 5) test a model that includes a general school climate factor (e.g., one-factor, second-order, or bifactor); and 6) are validated using samples predominantly comprised of English-speaking respondents from traditional K-12 schools in the U.S..

In order to ensure that school climate remains the focus of the study, only investigations that validated surveys measuring at least two core domains of school climate (as defined by: Wang & Degol, 2016) were included. Survey items were examined for face validity, and those that only assessed one domain of school climate (e.g., only physical environment or academic climate) were excluded. To ensure that the conceptualization of school climate as a distinct phenomenon was consistent across surveys, only studies that tested (though did not necessarily accept) a model that included a general school climate factor were included. Thus, articles that validated subdimensions of school climate separately, or that only tested a multifactor model and did not explore alternative factor structures that included a general school climate factor (e.g., unidimensional, second order, or bifactor structure), were excluded. In addition, because the measurement techniques to validate measures of lone individuals within schools differ than those used to validate measures of large groups of people, studies that report on measures of solely principals' or administrators' perceptions of school climate were excluded. In contrast to students, teachers, and parents, there is typically only one principal or a very small group of administrators within schools. Thus, there is no variability within schools.

The following types of research were considered beyond the scope of this review and were therefore excluded: 1) studies examining the school climate of classrooms, pre-schools, pre-kindergartens, higher education, or alternative/technical or other non-traditional schools; 2) measures that predominantly relate to only one core dimension of school climate (e.g. exclusively relational/connectedness or institutional environment dimensions); 3) measures that include non-survey formats (e.g. observation checklists or alternative indicators such as student-teacher ratio, attendance, and/or discipline records); and 4) studies for which survey development/validation was not a stated goal. In addition, studies in which the items used were not provided in full detail or could not be found elsewhere or obtained from the study's authors, were not included due to the inability to meaningfully assess content/face validity. If a study contained multiple measures or multiple studies only measures and studies that meet the above criteria were evaluated.

Selection Process

Identified articles were screened according to PRISMA guidelines (see Figure 1). Identified records from the database searches were imported into Endnote and de-duplicated. Screening and eligibility assessment were conducted in an un-blinded manner by the first author. Screening occurred in two stages. Initially, the titles were screened to exclude any studies that obviously violate the above criteria. The abstracts of remaining articles were screened in the same way. Any studies that potentially met the inclusion criteria were retrieved and the full text assessed for inclusion in the literature review synthesis.

Data Collection

Data Extraction

Articles were considered to contain multiple studies if they conducted analyses on separate participant samples and the full sample was never analyzed. Randomly split samples were not considered multiple studies. If an article contained multiple studies, they were analyzed separately. Information regarding study characteristics, validation strategies, and survey characteristics were collected.

Study Characteristics. For each study included in the synthesis, the following characteristics were collected and coded: "*Data Collection Year*," "*Participant Demographics*" (e.g. sample size, percentage of sample from various gender, racial/ethnic, and age/grade level groups), and "*School Demographics*" (e.g., school sample size; percentage of elementary, middle, and high schools; number of participants from each school, social/geographical information).

Validation Strategies. In addition to study characteristics, data regarding the survey validation strategies employed within each study were coded. For each study, "*Variance Across Levels*" information was collected (e.g., variability in items or factors attributed to the individual versus school level, such as intraclass correlation [ICC] coefficients). Additional methods of validation were organized by "*Data Level*" (e.g., were analyses conducted at the individual level or school level) and, further, by "*Approach to Nested Data*" (e.g., techniques employed, or not employed, to address the nested nature of the data, such as group-mean centering individual-level response data and using multilevel methods). "*Structural Validity Methods*" (e.g., exploratory factor analysis [EFA] and confirmatory factor analysis [CFA]), were coded, as were the estimation method utilized and the models (e.g., unidimensional, second order) tested.

"*Internal Consistency Methods*" (e.g., Cronbach's Alpha) were coded, including whether internal consistency was tested for each unidimensional scale. Validation methods that assessed the equality of the survey across diverse individual or school characteristics were coded under "*Invariance Strategies*." Information regarding the types of invariance analyses conducted (e.g., multiple-group CFA or multiple-indicators and multiple-causes [MIMIC] analysis) and the groups assessed was also collected. Methods of investigating a survey's associations to established measures of school climate or related constructs were coded under "*Construct Validity Analyses*." This section also included information about the type of analysis conducted (e.g., correlation) and the related constructs that were analyzed. The "*Subgroup Analyses*" section includes information regarding additional analyses conducted to explore group differences at the individual or school level). Finally, "*Other Validation Analyses*" was used to identify whether articles conducted factor correlations analyses.

Survey Characteristics. Finally, data regarding the characteristics of surveys were collected and synthesized. This included information about: "*Rater(s)*" (e.g. student, parent, or teacher/staff), "*Title or Author Description of School Climate Measure*", "*Target Age or School Level*" (e.g., author description of age/grade range or school type [elementary, middle, high] for which survey is appropriate), "*Conceptualization of Climate*" (e.g. individual- and/or school-level construct), "*Author Description of School Climate Construct*" (e.g., "school climate", "school connectedness", school experiences), "*Theory*" (e.g. the theoretical foundation(s), if any, used to guide measure development), "*Prior Validation Strategies*" (e.g., review of literature, expert panel, field tested), "*Purpose*" (e.g., author description of the purpose of the study, for example for DOE data, school practice, research), "*Dimensions Assessed*" (e.g., the presence of items targeting each school climate domain [according to descriptions outlined in Wang &

Degol, 2016]), "*Number of Items*" (i.e., the number of retained items), "*Item Referent(s)*" (e.g. whether items use the individual [e.g. "I treat other students fairly")] or school [e.g. "Students treat one another fairly"] as the reference), "*Number of Response Options*" (e.g., 4-point Likert scale), *"Model*" (e.g., the factor structure, such as first-order or second-order structures), "*Number of Factors*" (e.g., number of each type of factor), and "*Factors/Items*" (i.e., the number of retained items for each corresponding (sub)factor).

Collection Process

Two researchers reviewed the studies included in the subsequent synthesis and collected and coded information according to procedures recommended by Wilson (2009). Before coding began, the researchers discussed: the purpose of the synthesis and the data to be collected. An initial subset of studies was jointly coded to assure agreement regarding the use of the coding templates (described below). Following the joint coding effort, the two researchers independently coded a subset of studies and met to compare results. Disagreements were resolved by discussion and consensus.

Interrater reliability (IRR) refers to the percentage of coding that is the same between raters. IRR was calculated for each area of coding (i.e., Study Characteristics, Validation Strategies, and Survey Characteristics) with the intent of staying above 90%. The coding protocol was further discussed and operationalized as necessary. The final IRR was adequate in each category (Overall IRR = 99.6%; study characteristics IRR = 99.6% [range = 93.0% to 100% across studies]; validation strategies IRR = 99.9% [range = 99.3% to 100% across studies]; and survey characteristics IRR = 99.0% [range = 94.7% to 100% across studies]). Interrater reliability was less than 90% for only one variable, "*Conceptualization of Climate*" (IRR = 85.7%). Discrepancies in coding were discussed until consensus was reached.

Data Synthesis

Results were organized by raters (e.g. students, teachers/staff, parents). Validation strategy results were additionally organized by data level, and approach to nested data. Survey characteristic results were organized by raters and school climate measure; thus, results from studies validating the same measure were combined. Summary information for data items were presented when relevant (e.g., range, mean, and median). Per the goal of the current review, the descriptive synthesis of study results included the conceptualization of school climate and the validation strategies employed, with particular attention to any multilevel or invariance techniques employed. In addition, the absence of multilevel and measurement invariance techniques in analyses where they were necessary for valid interpretation was noted in the descriptive synthesis. For example, the appropriateness of construct validity analyses in the absence of multilevel techniques, or group difference analyses in the absence of measurement invariance, were discussed.

In addition, surveys were evaluated with respect to the methodological quality of reliability and validity analyses. Individual- and school-level validation strategies were assessed separately. The following domains were assessed: structural validity, internal reliability, measurement invariance, and construct validity, using a checklist based on the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN: Mokkink et al., 2017; Prinsen et al., 2018) (see Table 1.2). For each item in the checklist, studies were assessed on a 4-point scale: ‘good’, ‘adequate’, ‘doubtful’, and ‘inadequate’. For each area, the item “*Were there any other important flaws in the design or statistical methods of the study?*” was used to identify whether the analysis accounted for the nested nature of the data. Analyses that did not account for the nested nature of the data were rated as ‘doubtful’. Domains of reliability and validity were rated according to the ‘worst score counts’ method outlined in Mokkink et al. (2012). Thus, if one item in a domain was rated as ‘inadequate’ (e.g., the study did not investigate the internal consistency of each unidimensional factor), the quality of validation strategies for that domain was determined to be ‘inadequate’ quality. If multiple studies investigated the same survey, the study with higher ratings in a domain was used. An overall rating was calculated for each survey by assigning a “2” for ‘good’ domains, a “1” for ‘adequate’ and ‘doubtful’ domains, and “0” for ‘inadequate’ domains and domains that were not investigated. Thus, total scores could range from “0” (i.e., structural validity, internal reliability, measurement invariance, and construct validity methods were each rated as ‘inadequate’ or were not investigated) to “8” (i.e., structural validity, internal reliability, measurement invariance, and construct validity methods were each rated as ‘good’).

Risk of Bias

Like every systematic review, results may be biased across selected studies. Only articles published within peer reviewed journals were included within the study. Research with “negative” or “uninteresting” results may be less likely to be submitted for publication and/or accepted by journals (Gilbody & Song, 2000). Additional inclusion criteria may have also inadvertently biased results of the present review. For example, only articles published between 2007-2017 that validated surveys using samples predominantly comprised of participants from traditional K-12 schools in the U.S. and that tested a model that included an overall school climate factor were included. It is possible that school climate surveys that did not meet the inclusion criteria for the present article would demonstrate different trends. However, these criteria were specified in order to ensure that included surveys were of high quality and that conceptual and statistical trends could be meaningfully compared across studies. Potential effects of selection bias are discussed further in the *Limitations* section

Results

Study Selection

Figure 1 provides a visual representation of the search process. The search was conducted on March 11, 2018. Using the search terms: (“school climate” or “school environment” or “school atmosphere” or “school culture” or “educational climate” or “educational environment” or “educational atmosphere” or “educational culture”) and (“assess*” or “instrument*” or “inventory” or “measure*” or “scale” or “survey” or “test” or “tool” or “validation”) to locate articles published between 2007 and 2017 in Academic Search Complete, Education Source, ERIC, PsycARTICLES, Psychology and Behavioral Sciences Collection, and PsycINFO, 12,822 articles were identified. An additional 30 articles were identified through other sources (e.g., references from relevant articles). After, duplicates were removed, 8,521 records were initially screened to determine eligibility, with 6,857 articles excluded for reasons including: were not conducted in the United States, did not use data from traditional K-12 public schools, did not include quantitative analyses, were not empirical studies, did not use survey data. Twenty of the articles identified through other sources were also excluded because they were not published within the specified date range or were not peer-reviewed. After these records were removed, the remaining 1,634 records were screened a second time, with 1,072 excluded for similar reasons. The full texts of each of the remaining articles (n = 572) were then retrieved and assessed for eligibility. Of these, 558 were excluded for the following reasons: (a) study conducted no survey validation analyses or survey development/validation was not stated purpose (n = 414); (b) study did not investigate validity of an overall climate factor (e.g., only tested multifactor models, including EFAs; $n = 32$); (c) study validated a non-school climate measure ($n = 19$); (d) study assessed a specific type of climate and not general climate (e.g., racial climate or bullying climate, $n = 13$); (e) study was descriptive with no statistical analyses ($n = 12$); (f) only assessed one domain of climate (e.g., only the physical environment domain, $n = 14$); (g) study did not use data from traditional K-12 public schools ($n = 11$); (h) study did not include school climate survey data ($n = 11$); (i) study was theoretical and did not present a school climate survey ($n = 6$); (j) study was conducted outside of the United States ($n = 7$); (k) study was specific to a class or the classroom ($n = 5$); (l) study was a statistical primer and did not present a school climate survey (4); (m) school climate data was provided by a single rater (e.g., school administrator, $n = 4$); (n) school climate data were specific to a certain type of person or school (e.g., music teachers or military-connected schools, $n = 4$); and (o) study used participants from another country in combination with United States participants to validate school climate survey (n = 2). Several articles met exclusion criteria in multiple categories.

In total, 14 articles were included in the final synthesis. Three articles utilized multiple studies to investigate the presented survey (Anderson-Butcher et al., 2012; Bahena et al., 2016; Phillips & Rowley, 2015). All three of the studies from Bahena et al. (2016) and both of the studies from Phillips and Rowley (2015) were included in the synthesis. Only one study from Anderson-Butcher et al. (2012) was included, as the other study did not meet inclusion criteria (e.g., did not use data from traditional K-12 public schools). Within the 14 articles, 17 studies were analyzed.

Study Characteristics

Detailed information regarding study characteristics can be found in Table 1.3. Across the 17 studies, participant sample sizes ranged from 188 (Bahena et al., 2016, Study 2) to

500,800 (Furlong et al., 2011) (*M* sample size = 60,733; *Mdn* = 5,781). About half of the studies analyzed student surveys (*n* = 8), while the other half analyzed teacher/staff surveys (*n* = 4) or parent surveys (*n* = 5). Only a quarter of the studies validating student surveys (*n* = 2) utilized elementary students' perceptions; the majority utilized middle and/or high school students' participants (*n* =5). Except for Levitch et al. (2008), each of the teacher/staff and parent surveys were validated using participants from all school levels (elementary, middle, and high school).

Of the studies that provided gender or race/ethnicity demographics of student participants or participants' children, all utilized samples that were generally evenly split between males and females (*M* female = 50.4%; *M* male = 47.5%) and almost all (*n* = 10) utilized a majority White/European American sample (*M* = 53.3%). The two that did not (Furlong et al., 2011; Zullig et al., 2015) utilized a majority Hispanic/Latino sample. Furlong et al. (2011) further specified the 18 socio-cultural groups represented by their student participants– those groups were combined during synthesis for the purposes of comparison (e.g., students who identified as different Asian socio-cultural groups, such as Cambodian and Chinese, were combined to calculate the percent of participants who were Asian. Disaggregated socio-cultural demographics can be found below Table 1.3). Johnson et al. (2007) was the only study that provided gender or race/ethnicity demographics of teacher/staff or parent/guardian participants. They utilized a predominantly female (81%) teacher sample population, which is consistent with general characteristics of teachers. Most of their participants identified as White/European American (72%), with the majority of remaining participants identifying as Hispanic/Latino (24%).

School sample sizes were provided for about half of the studies (*n* = 8) and ranged from 7 to 1,073 (*M* = 230; *Mdn* = 110). Less than half of these (*n* = 3) provided information regarding how many participants per school were included. About half of the remaining studies (*n* = 5), specified that its participants were from multiple schools within a certain state or within specific school districts, and thus may have been affected by issues of participants being clustered within schools. The other half (*n* = 4) collected data from SurveyMonkey's national panel, and thus may not have been concerned with issues related to participants being clustered within schools (Bahena et al., 2016: Studies 1, 2, and 3; Schueler et al., 2014).

Almost all of the studies (*n* = 14) provided geographic information about their participants. About two-thirds (*n* = 10) of these were confined to a specific geographic region (i.e., a midwestern, northeastern, southeastern, southwestern, or western state), and about a third (*n* = 4) were conducted using a geographically diverse/national sample. About two-thirds of the studies (*n* = 11) specified when data were collected; the majority of which (*n* = 8) utilized data collected between 2011 and 2015. See Table 1.4 for a summary of study characteristics.

Validation Strategies

Detailed information regarding validation strategies can be found in Table 1.5. Of the 14 surveys validated, less than one-third (*n* = 4) were investigated to determine the degree to which variability in responses was accounted for at the individual versus school level (Bear et al., 2011, 2014; La Salle et al., 2016; Zullig et al., 2015). Results from those analyses showed that a meaningful amount of variability was due to school-level differences for all or some of the factors. Across surveys, nearly all validity analyses were conducted at the individual level.

Individual Level Analyses

Structural Validity. A little less than half of the 14 articles (*n* = 6) conducted exploratory factor analyses (EFAs), and all of the articles conducted confirmatory factor analyses (CFAs). About two-thirds of the articles (*n* = 9) tested multiple CFA models (e.g., one-factor, multi-factor, bifactor, second order) to determine the best fitting model. The remaining articles tested only a one-factor (*n* = 4) or second-order model (*n* = 1). Overall, more than three-quarters of the articles (*n* = 11) tested a one-factor model, about two-thirds (*n* = 9) tested a multi-factor model, a less than half tested a second order (*n* = 6) or bi-factor model (*n* = 4).

Internal Consistency. Almost three-quarters of the articles (*n* = 10) assessed the internal reliability of all survey factors. Two (Anderson-Butcher et al., 2012; You et al., 2014) only assessed the internal consistency of first-order factors, and two surveys were not investigated with respect to internal consistency (Levitch et al., 2008; Phillips & Rowley, 2015).

Invariance Analyses. More than three-quarters of the surveys (*n* = 11) were investigated to determine measurement invariance across various subgroups. Nearly all of these (*n* = 10) utilized multiple-group CFA (MGCFA) to examine measurement invariance. Schueler et al. (2014) utilized a multiple-indicators, multiple causes approach (MIMIC modeling). A little more than half of the articles (*n* = 8) investigated invariance across grade/school level groups, and about a third investigated invariance across gender (*n* = 5) or racial/ethnic (*n* = 5) groups. Half of the teacher/staff surveys were investigated with respect to invariance across different staff positions (Bear et al., 2014; You et al., 2014). Invariance across SES/income level, achievement level, and drop-out risk groups was tested least frequently (*n* = 2, 1, and 1, respectively).

Construct Validity. About a third of the articles (*n* = 5) analyzed construct validity at the individual level. Three student surveys were assessed to determine their relationships with school-level achievement (La Salle et al., 2016), school support (Furlong et al., 2011), or bullying/victimization (White et al., 2014) outcomes. Two parent surveys were investigated to determine their relationships with school climate, parent involvement, school satisfaction, and/or parent self-efficacy outcomes (Bahena et al., 2016; Schueler et al., 2014).

Means Analyses. Half of the articles (*n* = 7) analyzed means at the individual level. Most of these (*n* = 5) provided total scores or overall factor means. Two provided only item means.

Subgroup Analyses. A little less than half of the articles (*n* = 6) conducted individual-level subgroup analyses, all of which analyzed group differences in perceptions of school climate. Half of these (*n* =3) also analyzed group differences in associations with correlates: two included grouping variables in interaction coefficients within regression models (La Salle et al., 2016; White et al., 2014), and one conducted construct validity analyses separately for various subgroups (Furlong et al., 2011). Phillips and Rowley (2015) also conducted means analyses by subgroup but only to demonstrate that partial invariance had little impact on factor means. Thus, they were not coded as subgroup analyses.

School Level Analyses

Less than a third of the articles (*n* = 4) conducted analyses at the school level. Half of these (*n* = 2) analyzed means: one analyzed factor means by school-level subgroups (i.e., elementary and middle/high: Bear et al., 2011), and one utilized ANOVA to analyze mean differences across schools (Johnson et al., 2007). Three articles analyzed construct validity, investigating the association between school-level climate and indicators of school-level achievement, suspension/expulsion rates, or perceptions of bullying/victimization (Bear et al., 2011, 2014, 2015). Each of these also analyzed associations by school-level subgroups (e.g., elementary schools, middle/high schools). See Table 1.6 for a summary of validation strategies.

Survey Characteristics

Survey Purpose

Detailed information regarding survey characteristics can be found in Table 1.7. Half of the surveys (*n* = 7) assessed student perceptions, and the other half investigated teacher/staff (*n* =

4) or parent/guardian (*n* = 3) perceptions. Less than half of the student surveys (*n* = 3) specified the target school level (i.e., elementary, middle, and/or high) of the presented survey: one targeted all school levels (Bear et al., 2011); one targeted the elementary level (La Salle et al., 2016); and one targeted middle and high school students (Zullig et al., 2015). Except for Levitch et al. (2008), which did not specify a target school level, all of the teacher/staff and parent guardian surveys targeted all school levels. Almost all of the articles specified that their surveys were intended for practitioners (e.g., program evaluation, needs assessment, universal screeners) and/or researchers (*n* = 13 for each). Over half (*n* = 8) were also intended for state Department of Education use.

Survey Development

Nearly all of the articles (*n* = 12) specified the theoretical framework utilized to develop their survey. Authoritative discipline, bio-ecological, and risk and resilience theories were the most commonly referenced (*n* = 3 for each). Almost all of the articles (*n* = 12) referred to school climate literature in the development of their survey, and less than half referenced the use of previously validated surveys (*n* = 6), expert panels (*n* = 5), qualitative field-testing (*n* = 4), quantitative field-testing (*n* = 2), and/or stakeholder interviews (*n* = 2). Two articles did not discuss prior validation strategies (Levitch et al., 2008; Phillips & Rowley, 2015).

Conceptualization of Climate Construct

The majority of surveys (*n* = 11) referred to the measured construct as “school climate.” The remaining labeled the measured construct as “school experiences” (Anderson-Butcher et al., 2012), “school connectedness” (Furlong et al., 2011), or “school fit” (Bahena et al., 2016). Of these, Furlong et al. (2011) noted that “school connectedness” can be used as a proxy for school climate, while Bahena et al. (2016) distinguished “school fit” from school climate. Only about a third of the articles (*n* = 5) clearly articulated the level(s) at which their survey measured climate. Three of these conceptualized it at the individual level (Bahena et al., 2016; La Salle et al., 2016; You et al., 2014), and two conceptualized it at the school level (Bear et al., 2014, 2015).

Bahena et al. (2016) clearly differentiated “school fit” as an individual-level construct, distinct from “school climate”, which they conceptualized at the school level: “School climate refers to the overall ‘‘quality and character of school life’’ for all students. By contrast, school fit refers to how well that school environment… matches the needs of an individual student” (p. 122). La Salle et al. (2016) discussed the importance of measuring student-level climate: “Individuals’ perceptions, rather than others’ perspectives or some objective reality, are critical for understanding their behavior” (p. 56). You et al. (2014) indicated that school climate can be conceptualized across levels, “some argue that climate is a characteristic of the organization, while others contend that it is best thought of as a characteristic of the individual” (p. 153), and posited that “perhaps as a consequence of the lack of a coherent unit of theory, the unit of analysis for school climate is similarly muddled” (p.153). However, the authors stated that their study was “rooted in the conceptual lens that organizational climate reflects an individual’s personal perceptual experiences of a particular campus environment” (p. 168). Bear et al. (2014) described school climate as "the learning environment at the school-wide level” (p. 339), and Bear et al. (2015) noted that all of their items are designed with school as the referent, “as strongly recommended for measures of *school* climate (p. 118).

About two-thirds of the articles (*n* = 9) did not clearly articulate the level(s) at which their survey was intended to measure climate. Although, many discussed that school climate occurs at different levels. Furlong et al. (2011) wrote that school-level and individual-level factors “both equally influence” school connectedness (p. 988). Johnson et al. (2007) described school climate as “the psychosocial context in which teachers work and teach” (p. 834). Levitch et al. (2008) conceptualized school climate as “the combination of observable characteristics of a school, such as tangible resources, and intangible resources (e.g., shared beliefs and values amongst the school staff, students and community,” which impact interactions, values, and resources (p. 79). In addition, several articles stated that their survey could be utilized at the individual level or aggregated to the school level (e.g., Anderson-Butcher et al., 2012; Phillips & Rowley, 2015).

Survey Contents

Final survey length ranged from 5 (Furlong et al., 2011) to 42 (Zullig et al., 2015) items (*M* = 21, median = 15). About two-thirds of the surveys (*n* = 9) included both self- and school-referent items, less than a third (*n* = 4) included only school-referent items (Bear et al., 2014, 2015; Schueler et al., 2014; You et al., 2014), and only one included all self-referent items (Bahena et al., 2016). The majority of articles (*n* = 11) identified final survey models that included a general school climate factor, such as a unidimensional (all items load on to one, overall school climate factor, *n* = 6), second-order (items load on to specified first-order factors which load on to one second-order factor, *n* = 3), or bi-factor (items load on to various specific factors and a general school climate factor, *n* = 2) model. Anderson-Butcher et al. (2012) also determined that multifactor model resulted in acceptable fit. Only three surveys did not reflect a general school climate factor. The authors investigating these surveys determined that only a multi-factor model (i.e., items load on to specific factors that reflect sub-dimensions climate, but no overall factor is specified) best fit their survey data (Bear et al., 2014; Levitch et al., 2008; Zullig et al., 2015).

Survey items were analyzed to determine the dimensions of school climate (as specified by Wang & Degol, 2016) that were assessed. Over half of the surveys (*n* = 8) only assessed two dimensions of school climate; less than half assessed three or four dimensions of school climate (n = 3 for each). All of the surveys included items reflecting the *Community* domain, most often the *Quality of Relationships* dimension; and nearly all of the surveys (n = 12) included items targeting the *Safety* domain, most often the *Order and Discipline* dimension. Over half of the surveys (n = 8) included items targeting the *Academic* Domain, most often the *Teaching & Learning* dimension. Finally, less than a quarter of surveys (n = 3) included items targeting the *Institutional Environment* domain. See Table 1.8 for a summary of survey characteristics.

Synthesis of Results

A primary aim of the current systematic review was to identify whether the strategies used to validate surveys accounted for the clustered nature of school climate survey data. All structural validity analyses were conducted at the individual level. Only about a third of the articles (n = 5) utilized methods that accounted for the nested nature of the data when investigating structural validity. Nearly all of these (n = 4) centered raw response data around school means to control for school-level variability (Bear et al., 2011, 2014, 2015; Zullig et al., 2015), and one utilized robust standard errors to control for the nested nature of the data (Phillips & Rowley, 2015). The structural validity of nearly two-thirds of the surveys (n = 9) was assessed using raw response data, which does not account for the clustered nature of school climate data. Of these, the characteristics of two studies (Bahena et al., 2016; Schueler et al., 2014) may not have needed to account for nested data, as they were conducted using a national panel of online respondents. However, the issue of nested data was not discussed in these articles.

All internal consistency, measurement invariance, and factor correlation analyses were also conducted at the individual level. Zullig et al. (2015) was the only article that accounted for the nested nature of the data when investigating internal consistency or factor correlations. They utilized school-mean centered response data when assessing their CFA model. Of the 11 surveys for which measurement invariance was investigated, less than half (*n* = 5) addressed the clustered nature of the data by utilizing centered data (*n* = 4: Bear et al., 2011, 2014, 2015; Zullig et al., 2015) or robust standards errors (*n* =1: Phillips & Rowley, 2015).

Construct validity analyses were conducted at both the individual and school levels. Only one of the five articles that investigated it at the individual level addressed the issue of clustered data, but the issue was accounted for only partially. La Salle et al. (2016) investigated construct validity using hierarchical linear modeling (HLM) with perceptions of school climate score at level 1 and school-level correlates at level 2. However, perceptions of climate were calculated using a survey that was developed (i.e., was validated with respect to structural validity and internal consistency) using raw response data. Thus, while these analyses accounted for school-level variability in overall perceptions of climate, the survey used to calculate overall perceptions did not account for the nested nature of the data. None of the three articles that investigated construct validity at the school level accounted for the measurement or sampling error associated with nested data. Bear et al. (2011, 2014, 2015) aggregated factor scores from surveys that had been validated at the individual level to explore construct validity at the school level.

Half of the articles (*n* = 7) investigated climate means at the individual level, and two investigated it at the school level. None of these analyses accounted for the nested nature of the survey data, utilizing either raw response data (Anderson-Butcher et al., 2012; Bahena et al., 2016; Furlong et al., 2011; La Salle et al., 2016; Schueler et al., 2014; White et al., 2014; You et al., 2014) or aggregated data from surveys validated at the individual level (Bear et al., 2011; Johnson et al., 2007).

The present synthesis also aimed to determine whether the methods used to validate school climate surveys aligned with authors' conceptualizations of school climate and intended uses of the school climate surveys. None of the articles that clearly articulated the level at which their survey was intended to measure climate utilized structural validity methods that aligned with that conceptualization. The three surveys intended to measure individual-level climate were validated using raw response data, which confounds individual- and school-level effects (Bahena et al., 2016; La Salle et al., 2016; You et al., 2014); and the two surveys that intended to measure school-level climate were validated using school-mean centered response data at the individual level, which removes school-level effects (Bear et al., 2014, 2015). Over half of the surveys (n = 8) were intended to be used for statewide DOE initiatives, which suggests that results may likely be aggregated and analyzed at the school level. Structural validity analyses for each of these surveys were conducted at the individual level, using raw response data (Furlong et al., 2011; La Salle et al., 2016; White et al., 2014; You et al., 2014) or school-mean centered response data (Bear et al., 2011, 2014, 2015; Zullig et al., 2015). Finally, two studies stated that responses to their surveys, which were validated at the individual level, could be aggregated to create school level scores (Anderson-Butcher et al., 2012; Phillips & Rowley, 2015).

Another primary aim of the current systematic review was to identify whether the strategies used to validate school climate surveys investigated the invariance of surveys across diverse student and school populations. None of the surveys were investigated with respect to invariance across diverse school-level characteristics. Over three-quarters of the surveys (n = 11) were investigated with respect to invariance across individual-level characteristics (of which, only about half utilized methods that accounted for the clustering of individuals within schools). Invariance was most often investigated with respect to school/grade level, gender, and race/ethnicity subgroups. Of the six articles that investigated group differences in associations among survey responses and related variables, only one (Furlong et al., 2011) tested the invariance of the survey across subgroups to ensure results could be validly compared. Of the seven articles that analyzed group differences in perceptions of school climate, less than half (*n* = 3: Bahena et al., 2016; Schueler et al., 2014; You et al., 2014) tested the survey's invariance across each subgroup. Furlong et al. (2011) tested for invariance across subgroups of one demographic (race/ethnicity) but not across other demographics (gender and school/grade level).

Survey Quality

Each survey was assessed with respect to the methodological quality of reliability and validity analyses across four domains (i.e., structural validity, internal reliability, measurement invariance, and construct validity analyses). Individual- and school-level validation strategies were evaluated separately (see Table 1.2). For each survey, checklist items were assessed as 'good', 'adequate', 'doubtful', 'inadequate', or 'not applicable' (i.e., not analyzed). Then, each of the four domains of reliability and validity was rated according to the 'worst score counts' method (Mokkink et al., 2012). An overall score of methodological quality was also calculated for each survey. Domains of reliability and validity that were rated as 'good' were assigned a score of "2"; domains rated as 'adequate' or 'doubtful' were assigned a score of "1"; and domains rated as 'inadequate' or that were not investigated were assigned a score of "0". Scores for each of the four domains were aggregated to create a total score of methodological quality for each survey. Thus, total scores could range from "0" (i.e., structural validity, internal reliability, measurement invariance, and construct validity methods were rated as 'inadequate' or were not investigated) to "8" (i.e., each domain was rated as 'good').

Overall ratings of methodological quality for individual-level validation strategies ranged from 1 to 6 (see Tables 1.9 and 1.10). The *School Climate Measure* (*SCM*: Zullig et al., 2015) had the highest overall rating (total score = 6). The authors utilized school-mean centered data to account for the nested nature of the data when conducting structural validity, internal consistency, and measurement invariance analyses but did not conduct construct validity analyses. The *Delaware School Climate Surveys-Student*, *Teacher/Staff*, and *Home* (*DSCS-S*: Bear et al., 2011; *DSCS-T/S*: Bear et al., 2014; and *DSCS-H*: Bear et al., 2015) had the next highest ratings (total score = 5). For each survey, the authors utilized school-mean centered data to account for the nested nature of the data when conducting structural validity and measurement invariance analyses but did not account for the nested nature of the data when conducting internal consistency analyses and did not analyze construct validity at the individual level. The *Tripod School Climate Index* (Phillips & Rowley, 2015), *Add Health School Connectedness Scale* (*SCS*: Furlong et al., 2011), *Parent perceptions of school fit scale* (Bahena et al., 2016), and *Parent perceptions of school climate scale* (Schueler et al., 2014) each earned total scores of 4. Phillips & Rowley (2015) utilized robust standard errors to account for the nested nature of the data when conducting structural validity and measurement invariance analyses but did not conduct internal consistency or construct validity analyses. Furlong et al. (2011), Bahena et al. (2016), and Schueler et al. (2014) each analyzed structural validity, internal consistency, measurement invariance, and construct validity, but they did not account for the nested nature of the data. It is possible that Bahena et al. (2016) and Schueler et al. (2014) utilized samples that did not require the authors to account for clustered data. However, this was not discussed in the articles, and, thus, was not considered when assessing the surveys.

The remaining surveys earned overall ratings between 1 and 3 with respect to the methodological quality of individual-level validation strategies. The *Georgia Elementary School Climate Survey* (La Salle et al., 2016), *Georgia Brief School Climate Inventory* (*GaBSCI*: White et al., 2014), and *Revised version of the School Level Environment Questionnaire* (*Revised SLEQ*: Johnson et al., 2007) each earned a total score of 3. La Salle et al. (2016) and White et al. (2014) did not account for the nested nature of the survey data when conducting structural validity, internal consistency, and construct validity analyses and did not conduct measurement invariance analyses. While La Salle et al. (2016) utilized HLM when investigating construct validity, the authors used a survey that was developed using raw response data to calculate student perceptions of school climate. Johnson et al. (2007) did not account for the nested nature of the survey data when conducting structural validity, internal consistency, and measurement invariance analyses and did not investigate construct validity. The *Perceived School Experiences Scale* (*PSCS*: Anderson-Butcher et al., 2012) and *Brief, California School Climate Survey* (*Brief-CSCS*: You et al., 2014) each earned a rating of 2. The authors did not account for the nested nature of the survey data when conducting structural validity and measurement invariance analyses and did not conduct construct validity analyses. While the authors did investigate internal consistency, they did not assess the internal consistency of the total survey despite evidence supporting a second-order structure. The *Teacher Questionnaire* (Levitch et al., 2008) earned the lowest overall rating (total score = 1). Levitch et al. (2008) did not account for the nested nature of survey data when conducting structural validity analyses and did not investigate internal consistency, measurement invariance, or construct validity.

Overall ratings for school-level validation strategies ranged from 0 to 1 (see Tables 1.11 and 1.12). The majority of surveys were not validated at the school level. Only Bear et al. (2011, 2014, 2015) investigated the school-level survey in one of the four validity domains assessed within this synthesis. The *DSCS-S*, *DSCS-T/S*, and *DSCS-H* were each investigated at the school level with respect to construct validity. Each survey earned a score of '1' (doubtful), as they used aggregated response data that did not account for the nested nature of survey data, used aggregated scores from surveys that were validated at the individual level, and did not account for sampling error when aggregating scores.

Discussion

Subsequent to its inclusion in ESSA (2015), more and more states are reporting indicators of school climate (e.g., the School Climate Performance Indicator: CDE, 2019; the School Climate Star Rating index: GaDOE, 2015; the Strive HI School Climate measure: HIDOE, 2018) within their accountability systems (Education Commission of the States. 2018). Thus, accurately measuring school climate is more important than ever. High-quality measurement instruments are a prerequisite for ensuring that conclusions regarding important dimensions of climate, associations between school climate and various outcomes of interest, and differences in school climate across individuals and schools are trustworthy. However, conceptual and statistical challenges arise when attempting to measure school climate. While several reviews of school climate and its measurement have addressed conceptual issues and synthesized basic validation strategies (see, e.g., Kohl et al., 2013; Ramelow et al., 2015; Thapa et al., 2013; Wang & Degol, 2016; Zullig et al., 2010), the present study was designed to examine the conceptual and statistical trends within school climate measurement development, with particular focus on the use of strategies that, (1) account for the clustered nature of school climate survey data, and (2) determine the invariance of surveys across diverse student and school populations. Fourteen surveys were reviewed, each measuring student (n = 7), staff (n = 4), or guardian (n = 3) perceptions of school climate. Findings demonstrate that the strategies used to validate instruments often do not adequately address the statistical issues associated with school climate surveys and that confusion still exists regarding the conceptualization of school climate.

Statistical Issues

Multilevel Strategies

A primary aim of the present synthesis was to identify the frequency with which researchers use validation strategies that address the hierarchical nature of school climate data and the types of techniques employed. While previous reviews of school climate have stressed the need for validation strategies that address the nested nature of data (e.g., see Ramelow et al., 2015; Wang & Degol, 2016), the use of such strategies has yet to be synthesized in reviews of school climate surveys. Consistent with claims made by previous researchers (e.g., see Marsh et al., 2012; Konold et al., 2014; Schweig, 2014), the majority of school climate surveys reviewed in the present synthesis were not validated using strategies that addressed the multilevel nature of the data. Of the 14 selected articles, eight did not utilize any methods that accounted for clustered data and conducted all analyses using raw response data or aggregated response data. Such data conflates the individual-level school climate phenomenon with the school-level phenomenon, making it difficult to interpret the meaning of results. Individual- and school-level climate may have different dimensions, factor structures, and associations with variables (Konold et al., 2014; Marsh et al., 2009, 2012; Schweig, 2014). When these phenomena are conflated, researchers, educators, and policy makers may make incorrect conclusions about important components of school climate, school climate's relationship to outcomes of interest, and perceptions of school climate within and between schools.

Even among the six articles that did address the multilevel nature of school climate data, there is still a need for improvement. None of the articles utilized multilevel methods that would statistically control for both the measurement and sampling error associated with nested survey data and would allow for both individual- and school-level climate to be modeled and explored simultaneously (Marsh et al., 2009, 2012; P. Mehta & Neale, 2005; Muthén, 1991, 1994; Muthén & Asparouhov, 2011). Instead, all of the analyses that addressed the clustered nature of data occurred at the individual level. Centering response data around the school mean was the most common method utilized (4 articles). This method is only sufficient if the individual level is the sole level of interest, as it removes variability that results from the school level (Huang, 2016; Huang & Cornell, 2016b). However, three of the articles that utilized this method went on to investigate school-level climate using aggregate survey responses (Bear et al., 2011, 2014, 2015). Thus, mean-centering data was not sufficient, and the subsequent school-level analyses assumed the cross-level factorial invariance of the surveys and were subject to sampling error associated with the varying number of participants within each school (Morin et al., 2014).

One article (Phillips & Rowley, 2015) utilized robust standard errors to address the clustered nature of survey data. Robust standard errors correct for issues related to the data dependency of nested data but assumes cross-level factorial invariance (Kim et al., 2015). Phillips and Rowley (2015) did not test the assumption of cross-level invariance when investigating the *Tripod School Climate Index*; if cross-level invariance was violated, their analyses may have yielded biased fit statistics and parameter estimates (Kim et al., 2015; Wu & Kwok, 2012). The final article that addressed the clustered nature of survey data utilized HLM to investigate the relationship between school-level achievement and individual-level perceptions of school climate (La Salle et al., 2016). However, perceptions of school climate were calculated using scores from a survey that was developed using raw response data, which did not account for the clustering of students within school. Thus, while their HLM model controlled for the school-level effects of total school climate scores, the survey used to determine school climate scores was subject to measurement error, as results from structural validity and internal consistency analyses confounded individual- and school-level effects.

Most of the articles that utilized methods to address the multilevel nature of data inconsistently applied such methods. Only two articles utilized methods that accounted for the nested nature of survey data within each area of reliability and validity that the authors investigated: Phillips and Rowley (2015) utilized robust standard errors when investigating structural validity and measurement invariance, and Zullig et al. (2015) utilized school-mean centered data to investigate structural validity, internal consistency, and measurement invariance. Three articles only accounted for the nested nature of data when investigating structural validity and measurement invariance, but not when investigating internal consistency or construct validity (Bear et al., 2011, 2014, 2015). One article did not account for the clustering of students within schools when investigating the structural validity and internal consistency of the survey, but then accounted for the nested nature of total survey scores when investigating construct validity by utilizing HLM with overall perceptions of school climate at level 1 (La Salle et al., 2016).

While the present synthesis indicates that multilevel strategies are infrequently utilized in school climate measurement development, findings suggest that the use of such strategies are trending upward. Researchers seem to be increasingly heeding calls to utilize methods that account for the clustered nature of survey data (see, e.g., Konold et al., 2014; Marsh et al., 2009, 2012; Schweig, 2014). The six articles that addressed the issue of multilevel data (within at least one analysis) were all published during the latter half of years included in the present study. Four of the six were published during or after 2015. This suggests that school climate survey validation strategies may be trending toward addressing multilevel issues. This is further supported by findings from Ramelow et al.'s (2015) review of 12 school climate surveys published between 2003 and 2013. While not included in their synthesis, the authors mention in the discussion that only two surveys addressed the challenges of multilevel data by conducting further hierarchical analyses such as intraclass correlations. Thus, about 17% of the surveys reviewed by Ramelow et al. (2015) addressed the multilevel nature of survey data, while about 43% of the surveys reviewed in the present study addressed such issues.

Invariance Strategies

Another primary aim of the present synthesis was to identify the frequency with which researchers investigate the invariance of surveys across diverse student and school populations. For group comparisons to be valid, it is important to ensure that surveys measure school climate equally regardless of individual (e.g. gender or race/ethnicity) or school (e.g. economic resources or demographic makeup) factors. However, trends in the use of invariance testing have yet to be included in a review of school climate measures (Kohl et al., 2013; Ramelow et al., 2015; Zullig et al., 2010). Results from the present review demonstrate that school climate surveys are often investigated with respect to invariance across diverse individual-level groups but not across diverse school characteristics. Eleven of the 14 selected articles explored measurement invariance. All invariance analyses were conducted at the individual level. However, invariance in relation to school-level factors must be determined for comparisons across schools to be valid (Kim et al., 2012, 2015). If schools are rated in terms of their climate, as is increasingly occurring within state accountability systems (see, e.g., CDE, 2017; GaDOE, 2015), it is critical that the measures used not only work equally well across diverse student populations, but also across diverse school communities. Thus, there is a need for school climate researchers to establish the invariance of surveys with respect to school-level factors.

While the present synthesis demonstrates that most of the selected surveys were investigated to determine invariance across individual characteristics, findings suggest a need for improved strategies when conducting invariance analyses. Per ESSA (2015), states are required to disaggregate accountability data by student characteristics, including major racial and ethnic groups, family income, disability status, and language status. To ensure that comparisons and accountability inferences are accurate, measurement tools need to work equally well across subgroups. To ensure that measurement tools are working equally across subgroups, the nested nature of school climate survey data must be taken into account. If it is not accounted for, incorrect conclusions may be made (Kim et al., 2012, 2015). This is particularly pertinent for student characteristics that are often disproportionately clustered within schools, such as race/ethnicity and income (Frankenberg et al., 2003; Orfield & Frankenberg, 2014). For example, noninvariance in relation to student race/ethnicity may actually be attributable to school-level factors associated with schools of varying racial compositions, rather than a reflection of noninvariance with respect to individual students' race/ethnicity.

Of the 11 surveys for which individual-level invariance was explored, only five controlled for school-level effects when conducting invariance analyses (the *DSCS-S, DSCS-T/S, DSCS-H*, *SCM*, and *Tripod School Climate Index*). The invariance of four of these surveys (the *DSCS-S*, *DSCS-T/S*, *DSCS-H*, and *SCM*) was investigated using school-mean centered data, which, as discussed previously, is sufficient if the only level of interest is the individual level (Huang, 2016). However, three of these surveys (the *DSCS-S, DSCS-T/S,* and *DSCS-H*) were then utilized to investigate school-level climate. Thus, these analyses assumed the cross-level invariance of the surveys and its invariance with respect to diverse school-level characteristics. The invariance of the *Tripod School Climate Index* was investigated using robust standard errors, which assumes cross-level factorial invariance (Kim et al., 2015). If cross-level invariance was violated, results from these invariance analyses may have been biased (Kim et al., 2015; Wu & Kwok, 2012). Lastly, nine articles investigated group differences in perceptions of school climate and/or in relationships between school climate and outcomes of interest. Of these, only three tested the invariance of the survey across each group to ensure that results could be validly compared (Bahena et al., 2016; Schueler et al., 2014; You et al., 2014); however, they did not account for the nested nature of data when investigating invariance. Thus, there is a need for school climate researchers to increasingly adopt multilevel strategies when exploring the equality of their surveys across individual-level groups.

Conceptual Issues

Results of the current investigation suggest that confusion still exists regarding the conceptualization of school climate and the comparability of the school climate construct across surveys. Similar to Kohl et al. (2013), the present study found that most surveys included relational and organizational items. All of the 14 surveys included items reflecting the "Community" domain, and 12 included items reflecting the "Safety" domain (Wang & Degol., 2016). However, the "Academic" and "Institutional Environment" domains were inconsistently represented, making it difficult to compare findings across surveys. Only three surveys (the *SCM*, *Revised SLEQ*, and *Teacher Questionnaire*) included items across each of the four domains described by Wang & Degol (2016). Notably, the *SCM* and the *Teacher Questionnaire* were also two of only three surveys for which authors concluded that a multifactor model fit their data better than a model with an overall school climate factor (e.g., unidimensional, second-order, or bifactor model). This may suggest that some of the subdimensions less frequently included in school climate surveys (e.g., academic and institutional environment subdimensions) do not contribute to the distinct phenomenon of school climate and may, instead, be related constructs. This should be explored further in future research.

Regarding the multilevel conceptualization of school climate, the majority of articles did not clearly differentiate between individual- and school-level climate. Many articles described school climate as consisting of both affective and organizational characteristics but did not specifically conceptualize which level their survey measured. There is a need for school climate researchers to more precisely differentiate between individual- and school-level climate so that similarities and differences in the construct and its relationships with outcomes of interest can be discovered across levels. Bahena et al. (2016) conceptualized their construct at the individual level and described it as “school fit”, separate from but related to school-level climate. School climate research would benefit from making this distinction more ubiquitous. Labeling the individual-level climate phenomenon “school fit” and using “school climate” to refer to the school-level phenomenon may help make the conceptualization of the construct and the intended use of measures clear to both authors and readers. Thus, the methods needed to validate surveys could be identified easily and findings regarding the constructs would not be confused across levels.

Conclusion

The present study builds upon recent peer-reviewed syntheses of school climate measurement (see, e.g., Kohl et al., 2013; Ramelow et al., 2015; Zullig et al., 2010) by investigating trends in the use of advanced validation strategies that address the conceptual and statistical challenges associated with school climate survey data. Findings demonstrate that school climate measurement development often fails to adequately account for the nested nature of survey data and to appropriately align validation strategies with the conceptualization and intended use of the surveys. Nearly all validation strategies were conducted at the individual level, and the majority of selected articles did not account for the nested nature of survey data. While school climate is often discussed as a school level phenomenon (Ramelow et al., 2015), none of the included surveys were investigated with respect to structural validity, internal consistency, or measurement invariance at the school level. Despite this, four surveys were explored at the school level with respect to construct validity and mean differences, two articles stated that survey results could be aggregated to create school level scores, and eight of the surveys were intended to be used for statewide DOE initiatives, which suggests that results are likely to be aggregated and analyzed at the school level. The tendency to not address the hierarchical nature of survey data may be exacerbated by the confusion surrounding the conceptualization of school climate as an individual- or school-level construct. The majority of articles did not clearly articulate the level at which their survey was intended to measure school climate. Thus, comparing results from these surveys, without careful consideration of these conceptual and statistical inconsistencies, may lead to invalid conclusions.

When the individual- and school-level climate effects are confounded, researchers and educators may make incorrect conclusions about differences and similarities in the construct across diverse individuals and schools, about the important components of climate to target in school improvement initiatives, and about the relationships between climate and outcomes of interest. Such conclusions could have adverse policy implications. Based on the synthesis results, the present study recommends that future school climate research differentiates between the individual- and school-level climate constructs (e.g., “school fit” versus “school climate”) and that surveys clearly articulate which construct(s) they are intended to measure. This will make clear to researchers and consumers the intended use of the surveys and strategies needed to validate them. “School fit” surveys should be validated and tested for invariance at the individual level while controlling for school-level effects. “School climate” surveys should be validated and tested for invariance at the school level while controlling for individual-level effects and sampling error. Making this distinction ubiquitous within school climate research will ensure that the construct can be measured validly and that discovered results are trustworthy.

Limitations

Several limitations in this study should be acknowledged. A relatively small number of studies met inclusion criteria. A systematic review of published quantitative studies that reported on the original development or refinement of school climate surveys was conducted. Titles, abstracts, and full texts of the over 8,000 retrieved articles were thoroughly reviewed, and articles validating 14 different surveys met inclusion criteria. While this number may seem small, it is larger than the number of surveys included in recent reviews of school climate measures (e.g., n = 7, 12, and 5 for Kohl et al., 2013; Ramelow et al., 2015; Zullig et al., 2010, respectively). In addition, this finding was meaningful in that it indicates that further attention to the validation of school climate surveys may be needed.

Like all systematic reviews, results of the present study may be biased. Only articles published in refereed journals were included, creating a potential publication bias. However, this criterion was deemed necessary to ensure that surveys were accessible and had been reviewed for quality. Additional selection criteria may also have biased results. Three criteria in particular emerged as possible sources of bias during the selection process: studies needed to be validated using samples predominantly comprised of participants from traditional K-12 schools in the U.S.; the original development or refinement of a general school climate survey needed to be a stated focus of the study; and the article had to test the hypothesis that an overall school climate factor accounted for the data (e.g., through a unidimensional, second-order, or bifactor model). These criteria resulted in the exclusion of several studies. For example, only one study validating the *PSCS* within the Anderson-Butcher et al. (2012) article was included in the synthesis. The other study, which conducted construct validity analyses, was excluded because it used a sample of participants from an alternative school. In addition, several studies that conducted only single-item analyses, EFA, or PCA were not included, as they did not test the hypothesis of an overall school climate factor (see, e.g., Lohmeier & Lee, 2011; T. Mehta et al., 2013).

Most notably, some articles that have applied advanced multilevel strategies to validate school climate surveys did not meet inclusion criteria. For example, the *Authoritative School Climate Survey* has student and teacher versions and has been validated using advanced statistical and multilevel strategies (see, e.g., Huang et al., 2015; Huang, & Cornell, 2016a; Konold et al., 2014; Konold & Cornell, 2015a; 2015b; Konold & Shukla, 2017). However, the articles investigating these surveys did not test the hypothesis of an overall school climate factor and subscales were often validated separately. Therefore, the conceptualization of climate did not align with the present synthesis and were considered outside the scope of this study. Still, these surveys likely would have earned higher ratings of methodological quality than the surveys included in the present synthesis. For example, Konold and Cornell, (2015a) and Konold et al. (2014) utilized multilevel modeling approaches to account for the nested nature of the data when investigating the student version of the *Authoritative School Climate Survey*. These methods allowed for the authors to investigate individual- and school-level school climate constructs simultaneously, while controlling for measurement and sampling error associated with nested data. Structural validity, internal consistency, and construct validity were all investigated at both the individual- and school-levels using methods that accounted for the clustered nature of survey data. Thus, researchers interested in adopting multilevel strategies may find it helpful to review these articles, along with statistical primers that demonstrate the use of multilevel strategies (see, e.g., Huang, 2016; Huang & Cornell, 2016b; Jak, 2013; Marsh et al., 2009, 2012; Schweig, 2014). The one issue not addressed in the aforementioned articles that was discussed in the present synthesis is the invariance of surveys with respect to school-level factors. Readers interested in testing the invariance of surveys in relation to school-level factors are encouraged to review Jak (2013), Kim et al. (2012), and Kim and Cao (2015).

When conducting a systematic review, there are many subjective decisions that must be made. For example, what information to collect, which studies to include, and how to code subjective items of interest. The present study attempted to address these limitations by establishing interrater reliability and by being transparent when reporting the methods and when describing the aim of the synthesis: to identify trends in how surveys address the conceptual and statistical challenges associated with school climate measurement. The studies included in the present review are not an exhaustive list of the validation strategies being employed by school climate researchers. Instead, the demographic, conceptual, and statistical characteristics of the included studies allowed for meaningful comparisons of trends in school climate survey validation.

References

Adler, F. (1892). *The moral instruction of children* (Vol. 21). D. Appleton.

Anderson-Butcher, D., Amorose, A., Iachini, A., & Ball, A. (2012). The development of the Perceived School Experiences Scale. *Research on Social Work Practice, 22*(2), 186–194. http://dx.doi.org/10.1177/1049731511419866

Bahena, S., Schueler, B. E., McIntyre, J., & Gehlbach, H. (2016). Assessing parent perceptions of school fit: The development and measurement qualities of a survey scale. *Applied Developmental Science*, *20*(2), 121–134. http://dx.doi.org/10.1080/10888691.2015.1085308

Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2011). Delaware School Climate Survey-Student: Its factor structure, concurrent validity, and reliability. *Journal of School Psychology, 49,* 157–174. http://dx.doi.org/10.1016/j.jsp.2011.01.001

Bear, G., Yang, C., Harris, A, Mantz, L, Hearn, S., Boyer, D. (2016, November*). Technical manual for the Delaware School Survey: Scales of school climate; bullying victimization; student engagement; positive, punitive, and social emotional learning techniques; and social and emotional competencies*. Retrieved from http://wh1.oet.udel.edu/pbs/school-climate/administration-of-survey/

Bear, G. G., Yang, C., & Pasipanodya, E. (2015). Assessing school climate: Validation of a brief measure of the perceptions of parents. *Journal of Psychoeducational Assessment*, *33*(2), 115–129. http://dx.doi.org/10.1177/0734282914545748

Bear, G. G., Yang, C., Pell, M., & Gaskins, C. (2014). Validation of a brief measure of teachers' perceptions of school climate: relations to student achievement and suspensions.

Learning Environments Research, 17(3), 339–354. http://dx.doi.org/10.1007/s10984-014-9162-1

Bliese, P. D. (2000). Within-group agreement, nonindependence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist 32*(7), 513–531.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 1: Theoretical models of human development* (6th ed., pp. 793 – 828). New York: Wiley.

California Department of Education. (2019, May). *Local Performance Indicator Quick Guide*. Retrieved from https://www.cde.ca.gov/ta/ac/cm/documents/localindquickref2019.docx

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of personality and social psychology*, *95*(5), 1005–1018. http://dx.doi.org/10.1037/a0013193

Clifford, M., Menon, R., Gangi, T., Condon, C., & Hornung, K. (2012). *Measuring school climate for gauging principal performance: a review of the validity and reliability of publicly accessible measures*. Washington, DC: American Institutes for Research.

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

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York: Free Press.

Dunn, E. C., Masyn, K. E., Jones, S. M., Subramanian, S. V., & Koenen, K. C. (2015). measuring psychosocial environments using individual responses: An application of multilevel factor analysis to examining students in schools. *Prevention Science*, *16*(5), 718–733. http://dx.doi.org/10.1007/s11121-014-0523-x

Eccles, J. S., Lord, S. E., & Roeser, R. W. (1996). Round holes, square pegs, rocky roads, and sore feet: the impact of stage-environment fit on young adolescents' experiences in schools. In D. Cicchetti & S. Toth (Eds.), *Adolescence: opportunities and challenges* (pp. 47–92). Rochester: University of Rochester Press.

Education Commission of the States. (2018, May). *Accountability and Reporting: ESSA Plans*. Retrieved from http://ecs.force.com/mbdata/mbQuest5E?rep=SA172

Frankenberg, E., Lee C., & Orfield, G. (2003). *A Multiracial Society with Segregated Schools*. Cambridge, MA: Civil Rights Project, Harvard University.

Furlong, M. J., O'Brennan, L. M., & You, S. (2011). Psychometric properties of the Add Health School Connectedness Scale for 18 sociocultural groups. *Psychology in the Schools, 48*(10), 986–997. http://dx.doi.org/10.1002/pits.20609

Gangi, T. A. (2010). *School climate and faculty relationships: Choosing an effective assessment measure* (Doctoral dissertation, St. John's University New York).

Georgia Department of Education. (2015, May 4). *2014 School Climate Star Ratings data calculation guide for principals and district users*. Retrieved from http://www.gadoe.org/External-Affairs-and-Policy/Policy/Pages/School-Climate.aspx

Gilbody, S. M., Song, F. J. (2000). Publication bias and the integrity of psychiatry research. *Psychological Medicine, 30*(2), 253–258. http://dx.doi.org/10.1017/s0033291700001732

Goldhaber, D. E. (2000). *Theories of human development: Integrative perspectives*. Mountain View, CA: Mayfield Publishing.

Halpin, A. W., & Croft, D. B. (1963). The organizational climate of schools. University of Chicago: Midwest Administration Center.

Hawai'i Department of Education. (2018). *2018-19 Strive HI Measures and Calculations: Technical Guide*. Retrieved from http://www.hawaiipublicschools.org/DOE%20Forms/StriveHI2019/2019TechnicalGuide.pdf

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education*, *84*(1), 175–196. http://dx.doi.org/10.1080/00220973.2014.952397

Huang, F. L., & Cornell, D. G. (2016a). 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. http://dx.doi.org/10.1177/0734282915621439

Huang, F. L., & Cornell, D. G. (2016b). Using multilevel factor analysis with clustered data: Investigating the factor structure of the Positive Values Scale. *Journal of Psychoeducational Assessment*, *34*(1), 3–14. http://dx.doi.org/10.1177/0734282915570278

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. http://dx.doi.org/10.1111/josh.12340

Jak, S. (2013). *Cluster bias: Testing measurement invariance in multilevel data*.

Johnson, B., Stevens, J. J., & Zvoch, K. (2007). Teachers' perceptions of school climate: A validity study of scores from the Revised School Level Environment Questionnaire. Educational and Psychological Measurement, 67(5), 833–844. http://dx.doi.org/10.1177/0013164406299102

Kim, E. S., Kwok, O., & Yoon, M. (2012). Testing factorial invariance in multilevel data: A Monte Carlo study. *Structural Equation Modeling*, *19*, 250–267. http://dx.doi.org/10.1080/10705511.2012.659623

Kallestad, J. H. (2010). Changes in school climate in a long-term perspective. *Scandinavian Journal of Educational Research*, *54*(1), 1–14 http://dx.doi.org/10.1080/00313830903488429

Kim, E. S., Yoon, M., Wen, Y., Luo, W., & Kwok, O. M. (2015). Within-level group factorial invariance with multilevel data: Multilevel factor mixture and multilevel MIMIC models. Structural Equation Modeling: A Multidisciplinary Journal, 1–14. http://dx.doi.org/10.1080/10705511.2014.938217

Kohl, D., Recchia, S., & Steffgen, G. (2013). Measuring school climate: an overview of measurement scales. *Educational Research*, *55*(4), 411–426. http://dx.doi.org/10.1080/00131881.2013.844944

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

Konold, T., & Cornell, D. (2015b). Multilevel multitrait–multimethod latent analysis of structurally different and interchangeable raters of school climate. *Psychological Assessment*, *27*(3), 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*(3), 238–255. http://dx.doi.org/10.1037/spq0000062

Konold, T., Cornell, D., Shukla, K., & Huang, F. (2017). Racial/ethnic differences in perceptions of school climate and its association with student engagement and peer aggression. *Journal of Youth & Adolescence, 46*(6), 1289–1303. http://dx.doi.org/10.1007/s10964-016-0576-1

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*(1), 54–69. http://dx.doi.org/10.1080/10627197.2016.1271705

La Salle, T. P., Zabek, F., & Meyers, J. (2016). Elementary student perceptions of school climate and associations with individual and school factors. *School Psychology Forum, 10*(1), 55–65.

Levitch, A., Placier, P., Morrison, J., Pearce, T., & Lambert, M. (2008). Measurement of perceptions of school climate in a midwestern state. *The International Journal of Learning: Annual Review, 14*(11), 79–88. http://dx.doi.org/10.18848/1447-9494/cgp/v15i11/46012

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., & ... Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, *62*(10), e1-e34. http://dx.doi.org/10.1016/j.jclinepi.2009.06.006

Lohmeier, J. H., & Lee, S. W. (2011). A school connectedness scale for use with adolescents. *Educational Research and Evaluation, 17*(2), 85–95. http://dx.doi.org/10.1080/13803611.2011.597108

Loukas, A. (2007). What is school climate? *Leadership Compass, 5*(1), 1-3.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects. *Educational Psychologist, 47*(2), 106–124. http://dx.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. http://dx.doi.org/10.1080/00273170903333665

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*(3), 259–284. http://dx.doi.org/10.1037/1082-989X.10.3.259

Mehta, T. G., Atkins, M. S., & Frazier, S. L. (2013). The Organizational Health of Urban Elementary Schools: School Health and Teacher Functioning. *School Mental Health, 5*(3), 144–154. http://dx.doi.org/10.1007/s12310-012-9099-4

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2017). COSMIN Risk of Bias checklist for systematic reviews of Patient-

Reported Outcome Measures. *Quality of Life Research, 27*(5), 1171–1179. http://dx.doi.org/10.1007/s11136-017-1765-4

Mokkink, Lidwine B., Terwee, Caroline B., Patrick, Donald L., Alonso, Jordi, Stratford, Paul W., Knol, Dirk L., Bouter, Lex M., & de Vet, Henrica C. W. (2012). *COSMIN checklist manual*. Amsterdam: University Medical Center.

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

Muthén, B. O. (1991). multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*(4), 338–354. http://dx.doi.org/10.1111/j.1745-3984.1991.tb00363.x

Muthén, B.O. (1994). Multilevel covariance structure-analysis. *Sociological Methods & Research*, *22*, 376–398. http://dx.doi.org/10.1177/0049124194022003006

Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 15–40). New York, NY: Taylor & Francis.

Orfield, G., & Frankenberg, E. (2014). Increasingly Segregated and Unequal Schools as Courts Reverse Policy. Educational Administration Quarterly, 50(5), 718–734. http://dx.doi.org/10.1177/0013161x14548942

Perry. (1908). *The management of a city school*. New York: The Macmillan Company.

Phillips, S. F., & Rowley, J. F. S. (2015). The Tripod School Climate Index: An invariant measure of school safety and relationships. *Social Work Research, 40*(1), 31–39. http://dx.doi.org/10.1093/swr/svv036

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147–1157. http://dx.doi.org/10.1007/s11136-018-1798-3

Ramelow, D., Currie, D., & Felder-Puig, R. (2015). The assessment of school climate: Review and appraisal of published student-report measures. *Journal of Psychoeducational Assessment, 3*3(8), 731–743. http://dx.doi.org/10.1177/0734282915584852

Schueler, B. E., Capotosto, L., Bahena, S., McIntyre, J., & Gehlbach, H. (2014). Measuring parent perceptions of school climate. *Psychological Assessment, 26*(1), 314–320. http://dx.doi.org/10.1037/a0034830

Schweig, J. (2014). Cross-level measurement invariance in school and classroom environment surveys: Implications for policy and practice. *Educational Evaluation and Policy Analysis, 36*(3), 259–280. http://dx.doi.org/10.3102/0162373713509880

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

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

White, N., La Salle, T., Ashby, J. S., & Meyers, J. (2014). A brief measure of adolescent perceptions of school climate. *School Psychology Quarterly, 29(3),* 349–359. http://dx.doi.org/10.1037/spq0000075

Wilson, D. B. (2009). Systematic coding. In H. Cooper, Hedges, L. V., & Valentine, J. C. (Ed.), *The handbook of research synthesis and meta-analysis (2nd edition)* (pp. 159-176). New York: Russell Sage Foundation.

Wu, J.-Y., & Kwok, O. (2012). Using structural equation modeling to analyze complex survey data: A comparison between design-based single-level and model-based multilevel approaches. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(1), 16–35. http://dx.doi.org/10.1080/10705511.2012.634703

You, S., O'Malley, M. D., & Furlong, M. J. (2014). Preliminary development of the Brief–California School Climate Survey: Dimensionality and measurement invariance across teachers and administrators. *School Effectiveness and School Improvement, 25*(1), 153–173. http://dx.doi.org/10.1080/09243453.2013.784199

Zabek, F., Rice, K. G., Davis, D. E., La Salle, T. P., and Meyers, J. (2017). *A grateful climate: Measurement invariance, demographic differences, and the relationship between school climate and gratitude* [Unpublished manuscript]. Department of Counseling and Psychological Services, Georgia State University.

Zimmerman, M. A., & Arunkumar, R. (1994). Resiliency research: implications for schools and policy. *Social Policy Report, 8*, 1–18.

Zullig, K. J., Collins, R., Ghani, N., Hunter, A. A., Patton, J. M., Huebner, E. S., & Zhang, J. (2015). Preliminary development of a revised version of the School Climate Measure. *Psychological Assessment, 27*(3), 1072–1081. http://dx.doi.org/10.1037/pas0000070

Zullig, K. J., Koopman, T. M., Patton, J. M., & Ubbes, V. A. (2010). School climate: Historical review, instrument development, and school assessment. *Journal of Psychoeducational Assessment*, *28*(2), 139–152. http://dx.doi.org/10.1177/0734282909344205

Table 1.1

PRISMA Checklist and Corresponding Page Number in Present Study

Section/Topic	#	Checklist Item	Page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	N/A
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1-10
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	10
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	12-13
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	11
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	11
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	13
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	14, 16
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	14-16
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A*
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I_2) for each meta-analysis.	17-18
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	18-19
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	19-20
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	20-22
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A*
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	22-28
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	28-33
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	43-45
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	34-41
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	43-45
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	41-43
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	N/A

Table 1.2

Checklist for Methodological Quality Assessment

	Assessment				
Module/Item	Good	Adequate	Doubtful	Inadequate	N/A
Structural Validity					
Statistical Methods					
1. Was exploratory or confirmatory factor analysis performed?	Confirmatory factor analysis performed	Exploratory factor analysis performed		No exploratory or confirmatory factor analysis performed	N/A
2. Was the sample size included in the analysis adequate?	7 times the number of items and ≥100	At least 5 times the number of items and ≥100; OR at least 6 times number of items but <100	5 times the number of items but <100	< 5 times the number of items	
3. Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws: Nested data accounted for		Other minor methodological flaws: Did not account for nested data	Other important methodological flaws (e.g. inappropriate rotation method)	
Internal Consistency					
Design Requirements					
4. Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	Internal consistency statistic calculated for each unidimensional scale or subscale		Unclear whether scale or subscale is unidimensional	Internal consistency statistic NOT calculated for each unidimensional scale or subscale	
Statistical Methods					
5. Was Cronbach's alpha or Omega calculated?	Cronbach's alpha, or Omega calculated	Internal consistency method not clear.	Only item-total correlations calculated	No Cronbach's alpha or Omega and no item-total correlations calculated	N/A
6. Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws: Nested data accounted for		Other methodological flaws: Did not account for nested data	Additional important methodological flaws	
Measurement invariance					
7. Was an appropriate approach used to analyze the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	N/A
8. Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws: Nested data accounted for		Other methodological flaws: Did not account for nested data	Additional important methodological flaws	
Construct Validity					
Comparison with Outcome Measure					
9. Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	N/A
10. Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws: Nested data accounted for		Other methodological flaws: Did not account for nested data	Additional important methodological flaws	

Table 1.3

Coding Document of Study Characteristics

				Participant Demographics			School Demographics			
Rater/ Study	Study	Data Collection Year	Sample Size	Grade	Gender	Race/Ethnicity	Sample Size	School Level	Participants per School	Geographical Region/Other
Student										
Anderson-Butcher et al., 2012	1	*Not Specified*	773	7th = 8% 8th = 31% 9th = 9% 10th = 11% 11th = 9% 12th = 28% UnK = 3%	F = 51% M = 46% UnK = 3%	AA = 14% As = 1% EA = 70% H/L = 2% Multi = 9% UnK = 5%	*Not Specified*	Elem = *NS* Mid = *NS*	*Not Specified*	Schools from 2 Districts
Bear et al., 2011	1	2007	11,780	K-5th = 67% 6th-8th = 21% 9th–12th = 12%	F = 50% M = 50%	AA = 30% As = 3% EA = 51% H/L = 10% Other = 7%	85	Elem = 68% Mid = 20% High = 12%	*M* = 144 (range = 39-200)	- Delaware - Mean FRL = 44% - Represents approximately 50% of public schools in state
Furlong et al., 2011	1	2003-2004 & 2004-2005	500,800	7th = 36% 9th = 33% 11th = 28% Oth = 3%	F = 54% M = 46%	AA = 5% AI = 2% As† = 14% EA = 35% H/L† = 36% PI = 1% Multi = 7%	*Not Specified*	Mid = *NS* High = *NS*	*Not Specified*	California
La Salle et al., 2016	1	2013-2014	197,512	4th = 50% 5th = 50%	F = 50% M = 50%	AA = 36% As/PI = 4% EA = 42% H/L = 13% Other = 6%	1,073	Elem = 100%	*Not Specified*	- Georgia - 81% of public elementary schools in state
Phillips & Rowley, 2016	1 2	2012 2013	27,420 66,531	*Not Specified* 6th-8th = 41% 9th-12th = 49% UnK = 10%	*Not Specified* F = 45% M = 42% UnK = 13%	*Not Specified* AA = 15% As = 3% EA = 32% H/L = 15% UnK = 35%	101 222	*Not Specified* Mid = *NS* High = *NS*	*Not Specified* *Not Specified*	*Not Specified* *Not Specified*
White et al., 2014	1	*Not Specified*	130,968	6th = 49% 8th = 52%	F = 50% M = 50%	AA = 34% As = 3% EA = 49% H/L = 9% Other = 5%	*Not Specified*	Mid = 100%	*Not Specified*	- Georgia - Public Middle Schools
Zullig et al., 2015	1	*Not Specified*	1,643	9th = 22% 10th = 19% 11th = 41% 12th = 18%	F = 50% M = 50%	AA = 7% AI/AN = 8% EA = 19% H/L = 61% Other = 6%	7	High = 100%	*Not Specified*	- Arizona - Public High Schools
Teacher/ Staff										
Bear et al., 2014	1	2011	5,781	K-5th = 59% 6th-8th = 21% 9th-12th = 20%	Gender = *NS* T = 69% Ad = 3% CIS = 20% NIS = 8%	*Not Specified*	132	Elem = 65% Mid = 21% High = 14%	*Not Specified*	- Delaware - Sample represents 45% of public-school teachers in Delaware
Johnson et al., 2007	1	*Not Specified*	2,549	K-5th = 50% 6th-8th = 26% 9th-12th = 24%	F = 81% M = 19% T = 100%	AA = 1% AI = 2% As = 1% EA = 72% H/L = 24% Other = 4%	119	Elem = 67% Mid = 22% High = 11%	Range = 6-65	- Southwestern United States - Large, urban school district
Levitch et al., 2008	1	*Not Specified*	63,280	3rd–6th = *NS*	Gender = *NS* T(G) = 69% T(S) = 11% Ad. = 6% Coun. = 4% Lib. = 4% Other = 11%	*Not Specified*	*Not Specified*	Elem = *NS* Mid = *NS*	*Not Specified*	- Midwest - Statewide Questionnaire
You et al., 2014	1	2007-2008 or 2006-2007	4,800 *(Initial Sample 81,261)*	Elem = 33% Mid = 33% High = 33%	Gen = *NS* T=50% Ad=50%	*Not Specified*	*Not Specified*	Elem = *NS* Mid = *NS* High = *NS*	*Not Specified*	- California - Public schools from all 58 California counties
Parent/ Guardian										
Bahena et al., 2016	1	2012	323	PK-5th = 40% 6th-8th = 25% 9th-12th = 36% (*Children*)	Gender = *NS* Fa = 58% Mo = 36% Other = 6%	AA = 6% AI/AN = 1% As/PI = 4% EA = 75% H/L = 7% Multi/Oth = 8% (*Children*)	*Not Specified*	Elem = *NS* Mid = *NS* High = *NS*	*Not Specified*	- "Tremendous geographic diversity" (online panel) - Array of school types (public, private, charter)
	2	2012	188	PK-5th = 31% 6th-8th = 27% 9th-12th = 42% (*Children*)	Gender = *NS* Fa = 52% Mo = 40% Other = 8%	AA = 7% As/PI = 1% EA = 73% H/L = 9% Multi/Oth = 9% (*Children*)	*Not Specified*	Elem = *NS* Mid = *NS* High = *NS*	*Not Specified*	- "Tremendous geographic diversity" (online panel) - Array of school types (public, private, charter)

	3	2013	1,033	PK-5th = 42% 6th-8th = 24% 9th-12th = 34% (*Children*)	Gender = *NS* Fa = 45% Mo = 50% Other = 5%	AA = 6% AI/AN = 2% As/PI = 6% EA = 68% H/L = 8% Multi/Oth = 4% (*Children*)	*Not Specified*	Elem = *NS* Mid = *NS* High = *NS*	*Not Specified*	- "Tremendous geographic diversity" (online panel) - Array of school types (public, private, charter)
Bear et al., 2015	1	2013	16,173*	K-5th = 78% 6th-8th = 19% 9th-12th = 3% (*Children*)	F = 53% M = 46% (*Children*)	AA = 21% As = 5% EA = 50% H/L = 16% Multi/Oth = 7% (*Children*)	99	Elem = 75% Mid = 22% High = 3%	School enrollment = 232 to 1,530. Completion rate = 10% to 66%.	- Delaware - General Education Public Schools - Mean FRL = 58.7%
Schueler et al., 2014	1	*Not Specified*	904	PK-5th = 45% 6th-8th = 24% 9th-12th = 31% (*Participants' Children*)	Gender = *NS* Fa = 54% Mo = 40% Other = 6%	AI/AN = <0.5% AA = 6% As/PI = 3% EA = 75% H/L = 8% Multi/Oth = 8% (*Children*)	*Not Specified*	Elem = *NS* Mid = *NS* High = *NS*	*Not Specified*	- National Panel

Note: Percentages may not sum to 100% due to rounding or when the study's data did not sum to 100%. UnK = Unkown. NS = Not specified. F = Female. M = Male. AI = American Indian. AN = Alaska Native. AA = Black/African American. As = Asian. PI = Pacific Islander. EA = White/ European American. H/L = Hispanic/Latino. Multi = Multiracial. Oth = Other race/ethnicity. Ad = Administrator. CIS = Certified Instructional Staff. Coun = Counselor. Lib = Librarian. NIS = Non-Instructional Staff. T = Teacher. T(G) = General Education Teacher. T(S) = Special Education Teacher. Fa = Father. Mo = Mother. Other = Other guardian.

*Bear et al. (2015) demographic data was taken from the *Participants* section. Demographic data from the analyses section was slightly different.

†Furlong et al. (2011) demographic data was further disaggregated into 18 sociocultural groups, which were combined for comparison purposes within this study. Here, 14% Asian reflects: Asian Indian = 1%, Cambodian = 1%, Chinese = 4%, Filipino = 5%, Japanese = 1%, Korean = 2%, Laotian = 1%, Vietnamese = 2%; and 36% Hispanic/Latino reflects: Central American = 3% South American = 1%, Cuban = <0.5%, Mexican = 31%, Puerto Rican = 1%.

Table 1.4

Study Characteristics Summary

Study	Student								Teacher/Staff				Parent/Guardian				
Characteristics	AB	B11	Fur	LaS	Phi		Whi	Zul	B14	Joh	Lev	You	Bah			B15	Sch
					1	2							1	2	3		
Participant Sample Size																	
<500													X	X			
500-999	X																X
1,000-9,999								X	X	X		X			X		
10,000-99,999		X			X	X					X					X	
≥100,000			X	X			X										
Participant (or child's) School/Grade Level																	
Elementary		67		100					59	50	X	33	40	31	42	78	45
4th				*50*													
5th				*50*													
Middle	39	21	36			41	100		21	26	X	33	25	27	24	19	24
6th							*49*										
7th	*8*		*36*														
8th	*31*						*52*										
High	58	12	64			49		100	20	24		33	36	42	34	3	31
9th	*9*		*33*					*22*									
10th	*11*							*19*									
11th	*9*		*28*					*41*									
12th	*28*							*18*									
UK/NS	3				X	10											
Participant (or child's) Gender																	
Female	51	50	54	50		45	50	50		81						53	
Male	46	50	46	50		42	50	50		19						46	
Not Specified	3				X	13			X		X	X	X	X	X		X
Participant (or child's) Race/Ethnicity																	
Am Ind/Al Nat			2					8		2			1		2		<1
Afr Am	14	30	5	36		15	34	7		1			6	7	6	21	6
As	1	3	14	4		3	3			1			4	1	6	5	3
Eur Am	70	51	35	42		32	49	19		72			75	73	68	50	75
His/Lat	2	10	36	13		15	9	61		24			7	9	8	16	8
Multi/Other	9	7	8	6			5	6		4			8	9	4	7	8
UK/NS	5				X	35			X		X	X					
Staff Role																	
Teacher									69	100	80	50					
Administrator									3		6	50					
Other Staff									28		19						
Guardian Role																	
Father													58	52	45		54
Mother													36	40	50		40
Other Guardian													6	8	5		6
UK/NS																X	
School Sample Size																	
<100		X						X								X	
101-500					X	X			X	X							
>500				X													
UK/NS	X		X				X				X	X	X	X	X		X
Geographic Region																	
Midwest											X						
Northeast		DE							DE							DE	
Southeast				GA			GA										
Southwest								AZ		X							
West			CA									CA					
National													X	X	X		X
UK/NS	X				X	X											
Year Data was Collected																	
2000-2005			X														
2006-2010		X										X					
2011-2015				X	X	X			X				X	X	X	X	
UK/NS	X						X	X		X	X						X

Note: UK/NS = Unknown/Not Specified. Am Ind = American Indian. Al Nat = Alaska Native. Afr Am = Black/African American. As = Asian. Eur Am = White/European American. His/Lat = Hispanic/Latino. Multi/Other = Multiracial or Other race/ethnicity.

Table 1.5

Coding Document of Validation Strategies

Rater/ Study	Study	ICC/ Variance Across Levels	Data Level	Approach to Nested Data	Structural Validity EFA	CFA Est.	 Models Tested	Internal Consistency Method	 Factors Tested	Invariance Analyses Method	 Groups	Construct Validity Method	 Constructs	Means Analyses	Subgroup Analyses Means/ Perceptions	 Assoc.	Other Analyses
Student																	
Anderson-Butcher et al., 2012	N/A		Ind.	Raw response data	- PAF	- MLR	- Multifactor - Second Order	Alpha	- Sub-factors	MGCFA (*multifactor model*)	- Gender			- Overall - Sub-factors			Factor Correlations
Bear et al., 2011	N/A	Factor ICCs (range = .04 to .18)	Ind.	Raw response data				Alpha	- All Factors								Factor Correlations
				Centered data		- FIML	- One-Factor - Multifactor - Bifactor			MGCFA	- Gender - Race/Ethnicity - School Level						
			Sch.	Aggregated data								Correlation	- School-Level ELA Achievement - School-Level Math Achievement - School-Level Suspension/ Expulsion Rates	- Overall - Sub-factors	- School Level (Elem, Mid/High)	- School Level (Elem, Mid/High)	
Furlong et al., 2011	N/A		Ind.	Raw response data		- Robust Estimation	- One-Factor	Alpha	- All Factors	MGCFA	- Race/Ethnicity	Correlation	- School Support Scale	- Overall	- Gender - Race/Ethnicity (18 groups) - Grade Level (7^{th}, 9^{th}, 11^{th})	- Race/Ethnicity (18 groups)	
La Salle et al., 2016	N/A	Survey variance across levels (7% of variance at school level)	Ind.	Raw response data		*NS*	- One-Factor	*NS*	- All Factors					- Overall	- Gender - Race/Ethnicity (EA, Minority) - Grade Level (4^{th}, 5^{th})		
				HLM with survey at level 1								Regression	- School-Level College and Career Readiness Performance Index			*Interactions* - Gender - Race/Ethnicity (EA, Minority) - Grade Level (4^{th}, 5^{th})	
Phillips & Rowley, 2016	1		Ind.	Robust Standard Errors	- WLS												
	2		Ind.	Robust Standard Errors		- WLS - ML	- One-Factor			MGCFA	- Gender - Race/Ethnicity (AA, EA, H/L) - School Level (Mid, High) - Achievement (A, B, C-F ranges) - SES (Low, Mod, High) - Drop-Out Risk (No, Some-High)						
White et al., 2014	N/A		Ind.	Raw response data	*NS*	*NS*	- One-Factor	Alpha	- All Factors			Regression	- Likelihood of helping someone else being bullied - Prevalence of	- Overall	- Gender - Race/Ethnicity (As, AA, H/L, EA, Other)	*Interactions* - Gender - Race/Ethnicity	

													bullying others, being bullied, and being teased		- Grade Level (6th, 8th)	(As, AA, H/L, EA, Other) - Grade Level (6th, 8th)	
Zullig et al., 2015	N/A	Factor ICCs (range = .00 to .14)	Ind.	Raw response data	*NS*			Alpha	- All Factors								
				Centered data		*NS*	- Multifactor - Second Order - Bifactor	Alpha	- All Factors	MGCFA	- Gender - Race/Ethnicity (H/L, EA)						Factor Correlations
<u>Teacher/Staff</u>																	
Bear et al., 2014	N/A	Factor ICCs (range = .02 to .29)	Ind.	Raw Response Data				Alpha	- All Factors								Factor Correlations
			Ind.	Centered data		- FIML	- One-Factor - Multifactor - Second Order - Bifactor			MGCFA	- School Level (Elem, Mid, High) - Staff Role (T, CIS, NIS)						
			Sch.	Aggregated data								Correlation	- School-Level ELA Achievement - School-Level Math Achievement - School-Level Suspension/ Expulsion Rates			- School Level (Elem, Mid, High)	
Johnson et al., 2007	N/A		Ind.	Raw response data	- PAF	*NS*	- Second Order	Alpha	- All Factors	MGCFA	- School Level (Elem, Mid, High)						
			Sch.	Aggregated Raw Data										*ANOVA* - Overall - Sub-factors			
Levitch et al., 2008	N/A		Ind.	Raw response data	- PAF	*NS*	- One-Factor - Multifactor - "Others"										Factor Correlations
You et al., 2014	N/A		Ind.	Raw Response Data		*NS*	- One-Factor - Multifactor - Second Order	Alpha	- Sub-factors	MGCFA	- School Level (Elem, Mid, High) - Staff Role (Ad, T)			- Sub-factors	- School Level (Elem, Mid, High) - Staff Role (Ad, T)		
<u>Parent/Guardian</u>																	
Bahena et al., 2016	1		Ind.	Raw response data		- WLSMV	- One-Factor - Multifactor	Alpha	- All Factors					- Items			
	2		Ind.	Raw response data		- WLSMV	- One-Factor	Alpha	- All Factors			Correlation	- School Satisfaction - School Climate - Parent Self-Efficacy				
	3		Ind.	Raw response data		- WLSMV	- One-Factor			MGCFA	- School Level (Elem, Mid, High) - Income Level (Lower, Higher)			- Overall	- School Level (Elem, Mid, High) - Income Level (Lower, Higher)		
Bear et al., 2015	N/A		Ind.	Raw response data				Alpha	- All Factors								- Factor Correlations
			Ind.	Centered data		- FIML	- One-Factor - Multifactor - Second Order			MGCFA	- Child's Gender - Child's Race/						

							- Bifactor				Ethnicity (AA, As, EA, H/L) - School Level (Elem, Mid, High)						
			Sch.	Aggregated data								Correlation	- School-Level ELA Achievement - School-Level Math Achievement - School-Level Suspension/ Expulsion Rates - School-Level Bullying/ Victimization Prevalence			- School Level (Elem, Mid)	
Schueler et al., 2014	N/A		Ind.	Raw response data		- WLSMV	- One-Factor - Multifactor	Alpha	- All Factors	MIMIC	- School Level (Elem, Mid/High)	Correlation	- School Satisfaction - Parent Self-Efficacy - Parent Involvement	- Items	*Path Analysis* - School Level (Elem, Mid/High)		

Table 1.6

Validation Strategies Summary

Validation Strategies	Student							Teacher/Staff					Parent/Guardian				
	AB	B11	Fur	LaS	Phi		Whi	Zul	B14	Joh	Lev	You		Bah		B15	Sch
					1	2							1	2	3		
ICC/Variance Across Levels		X		X				X	X								
Individual Level Analyses																	
Structural Validity																	
EFA (raw data)	X						X	X		X	X						
EFA (robust *SE*s)					X												
CFA (raw data)	X		X	X			X			X	X	X	X	X	X		X
CFA (centered data)		X						X	X							X	
CFA (robust *SE*s)						X											
Models Tested																	
One-Factor		X	X	X		X	X		X		X	X	X	X	X	X	X
Multi-Factor	X	X						X	X		X	X	X			X	X
Second Order	X							X	X	X		X				X	
Bi-Factor		X						X	X							X	
Internal Consistency																	
All Unidimensional Factors		X	X	X			X	X	X	X			X	X		X	X
Only (Sub)factors	X											X					
Invariance																	
MGCFA (raw data)	X		X							X		X			X		
MIMIC (raw data)																	X
MGCFA (centered data)		X						X	X							X	
MGCFA (robust *SE*s)						X											
Groups Assessed																	
Gender	X*	X				X		X								X*	
Racial Ethnicity		X	X			X		X								X*	
School/Grade Level		X				X			X	X		X			X*	X*	X*
Achievement						X											
SES/Income						X									X		
Drop-Out Risk						X											
Position (Staff)									X			X					
Construct Validity																	
Achievement (School Level)				X													
School Support			X														
School Satisfaction														X			X
School Climate														X			
Parent Self-Efficacy														X			X
Bullying/Victimization							X										
Parent Involvement																	X
Other Validation Analyses																	
Correlation Among Factors	X	X						X	X		X					X	
Means Analyses																	
Item Means													X				X
Overall Means	X		X	X			X					X			X		
Sub-Factor Means	X											X					
Subgroup Analyses																	
Means/Perceptions																	
Gender			X	X			X										
Racial/Ethnic			X	X			X										
School/Grade Level			X	X			X					X			X		X*
SES/Income															X		
Position (Staff)												X					
Associations with Correlates																	
Gender				X			X										
Racial/Ethnic			X	X			X										
School/Grade Level				X			X										
School Level Analyses																	
Construct Validity																	
Achievement		X							X							X	
Suspension/Expulsion		X							X							X	
Bullying/Victimization																X	
Means Analyses		X								X							
Subgroup Analyses		X							X							X	

.

Table 1.7

Coding Document of Survey Characteristics

Rater/ Study	Measure	Target School Level	Concept. of Climate	School Climate Construct	Theory	Prior Validation Strategies	Purpose	Dimensions Assessed (Final)				# of Items	Item Refer.	# of Resp. Opt.	Model/s	# of Factors	Factors/ Items
								Academic	Community	Institutional Environment	Safety						
Student																	
Anderson-Butcher et al., 2012	Perceived School Experiences Scale (PSES)	*NS*	Not Clear	School Experiences	- Risk & Resilience	- Review of Literature	- Practice	- Teaching & Learning	- Connectedness - Quality of Relationships			14	- Self - School	5	- Multi-Factor - Second-Order	1 Second Order 3 First Order	- Academic Press (4) - Academic Motivation (6) - School Connectedness (4)
Bear et al., 2011	Delaware School Climate Survey-Student (DSCS-S)	- Elem - Mid - High	Not Clear	School Climate	- Authoritative Discipline - Bio-ecological - Social Cognitive - Stockard and Mayberry's framework of school climate	- Review of Literature - Expert Panel - Field Tested Qualitative - Field Tested Quantitative	- DOE - Practice - Research		- Connectedness - Quality of Relationships		- Order & Discipline - Physical - Social/ Emotional	23	- Self - School	4	Bifactor	1 General 5 Specific	- Teacher-Student Relations (8) - Student-Student Relations (4) - Fairness of Rules (4) - School Safety (3) - Liking of School (4)
Furlong et al., 2011	Add Health School Connectedness Scale (SCS)	*NS*	Not Clear	School Connectedness	- Risk & Resilience	- Review of Literature - Previously Validated Survey	- DOE - Practice - Research		- Connectedness - Quality of Relationships		- Physical	5	- Self - School	5	Unidimensional	1	- School Connectedness (5)
La Salle et al., 2016	Georgia Elementary School Climate Survey	- Elem (3 - 5)	Individual	School Climate	- Bio-ecological	- Review of Literature - Previously Validated Survey	- DOE - Practice - Research		- Connectedness - Quality of Relationships		- Order & Discipline - Physical - Social/ Emotional	11	- Self - School	*NS*	Unidimensional	1	- School Climate (11)
Phillips & Rowley, 2015	Tripod School Climate Index	*NS*	Not Clear	School Climate	*NS*	*NS*	- Practice - Research		- Quality of Relationships		- Order & Discipline - Physical	7	- Self - School	5	Unidimensional	1	- School Climate (7)
White et al., 2014	Georgia Brief School Climate Inventory (GaBSCI)	*NS*	Not Clear	School Climate	*NS*	- Review of Literature	- DOE - Practice - Research	- Teaching & Learning	- Connectedness - Quality of Relationships		- Order & Discipline - Physical - Social/ Emotional	9	- Self - School	2, 3, 4	Unidimensional	1	- School Climate (9)
Zullig et al., 2015	School Climate Measure (SCM)	- Mid - High	Not Clear	School Climate	- Bio-ecological - Organizational Change Theory	- Review of Literature - Expert Panel - Previously Validated Surveys	- DOE - Practice - Research	- Teaching & Learning	- Connectedness - Partnership - Quality of Relationships - Respect for Diversity	- Environmental	- Order & Discipline - Social/ Emotional	42	- Self - School	5	Multi-Factor	10	- Positive Student-Teacher Relationships (8) - Order & Discipline (6) - Opportunities for Engagement (6) - School Physical Environment (4) - Academic Support (4) - Parent Involvement (3) - School Connectedness (4) - Perceived Exclusion/ Privilege (3) - School Social Environment (2) - Academic Satisfaction (2)
Teacher/Staff																	

Bear et al., 2014	Delaware School Climate Survey–Teacher/Staff (DSCS–T/S)	- Elem - Mid - High	School	School Climate	- Authoritative Discipline - Stockard and Mayberry's theoretical framework of school climate	- Review of Literature - Previously Validated Survey	- DOE - Practice - Research		- Partnership - Quality of Relationships - Respect for Diversity		- Order & Discipline - Physical - Social/ Emotional	24	- School	4	Multi-Factor	7	- Teacher-Student Relations (3) - Student-Student Relations (4) - Teacher-Home Communication (5) - School Safety (3) - Clarity of Expectations (3) - Fairness of Rules (3) - Respect for Diversity (3)
Johnson et al., 2007	Revised School Level Environment Questionnaire (Revised SLEQ)	- Elem - Mid - High	Not Clear	School Climate	- Moos's 1974 general categories of environments	- Review of Literature - Previously Validated Survey	- Practice - Research	- Leadership - Teaching & Learning	- Quality of Relationships - Respect for Diversity	- Availability of Resources	- Order & Discipline	21	- Self - School	*NS*	Second Order	1 Second Order 5 First Order	- Collaboration (6) - Decision Making (3) - Instructional Innovation (4) - Student Relations (4) - School Resources (4)
Levitch et al., 2008	Teacher Questionnaire	*NS*	Not Clear	School Climate	- Risk & Resilience	*NS*	- Research	- Leadership - Professional Development - Teaching & Learning	- Connectedness - Partnership - Quality of Relationships - Respect for Diversity	- Availability of Resources - Environmental - Structural Organization	- Order & Discipline - Physical - Social/ Emotional	81	- Self - School	5	Multi-Factor	2	- Intrinsic Learning Climate (48) - Extrinsic Learning Climate (33)
You et al., 2014	Brief–California School Climate Survey (Brief-CSCS)	- Elem - Mid - High	Individual	School Climate	- Moo's Organizational Climate Theory	- Review of Literature - Previously Validated Survey	- DOE - Practice - Research	- Teaching & Learning	- Connectedness - Partnership - Quality of Relationships		- Order & Discipline - Social/ Emotional	15	- School	*NS*	Second Order	1 Second Order 2 First Order	- Organizational supports (7) - Relational supports (8)
<u>Parent/Guardian</u>																	
Bahena et al., 2016	Parent perceptions of school fit scale	- Elem - Mid - High	Individual	School Fit	- Stage-Environment Fit	- Review of Literature - Expert Panel - Stakeholder Interviews - Field Tested Qualitative	- Practice - Research	- Teaching & Learning	- Connectedness - Quality of Relationships - Respect for Diversity		- Order & Discipline	7	- Self	5	Unidimen-sional	1	- School Fit (7)
Bear et al., 2015	Delaware School Climate Survey-Home (DSCS-H)	- Elem - Mid - High	School	School Climate	- Authoritative Discipline - Stockard and Mayberry's framework of school climate	- Review of Literature - Expert Panel - Field Tested Qualitative - Field Tested Quantitative - Previously Validated Surveys	- DOE - Practice - Research		- Partnership - Quality of Relationships - Respect for Diversity		- Order & Discipline - Physical - Social/ Emotional	26	- School	4	Bifactor	1 General 7 Specific	- Teacher-Student Relations (4) - Student-Student Relations (4) - Teacher-Home Communication (4) - School Safety (3) - Clarity of Expectations (4) - Fairness of Rules (4) - Respect for Diversity (3)
Schueler et al., 2014	Parent perceptions of school climate scale	- Elem - Mid - High	Not Clear	School Climate	- Stage-Environment Fit	- Review of Literature - Expert Panel - Stakeholder Interviews - Field Tested Qualitative	- Practice - Research	- Leadership - Teaching & Learning	- Connectedness - Quality of Relationships - Respect for Diversity			7	- School	5, 7	Unidimen-sional	1	- School Climate (7)

Note: Practice refers to practice purposes for school personnel and/or program evaluation.

Table 1.8

Summary of Survey Characteristics

Survey Characteristics	Student							Teacher/Staff				Parent/Guardian		
	AB	B11	Fur	LaS	Phi	Whi	Zul	B14	Joh	Lev	You	Bah	B15	Sch
<u>Target School Level</u>														
Elementary		X		X				X	X		X	X	X	X
Middle		X					X	X	X		X	X	X	X
High		X					X	X	X		X	X	X	X
Not Specified	X		X		X	X				X				
<u>Conceptualization of Level</u>														
Individual				X							X	X		
School								X					X	
Not Clear	X	X	X		X	X	X		X	X				X
<u>School Climate Construct</u>														
School Climate		X		X	X	X	X	X	X	X	X		X	X
School Connectedness			X											
School Experiences	X													
School Fit												X		
<u>Theoretical Model</u>														
Authoritative Discipline		X						X					X	
Bio-ecological		X		X			X							
Risk & Resilience	X		X							X				
Social Cognitive		X												
Stage-Environment Fit												X		X
Other		X					X	X	X		X		X	
Not Specified					X	X								
<u>Prior Validation Strategies</u>														
Review of Literature	X	X	X	X		X	X	X	X		X	X	X	X
Expert Panel		X					X					X	X	X
Stakeholder Interviews												X		X
Field Tested: Qualitative		X										X	X	X
Field Tested: Quantitative		X											X	
Previously Validated Survey			X	X			X	X	X		X		X	
Not Specified					X					X				
<u>Purpose</u>														
DOE Data		X	X	X		X	X	X			X		X	
Practice	X	X	X	X	X	X	X	X	X		X	X	X	X
Research		X	X	X	X	X	X	X	X	X	X	X	X	X
<u>Item Referent</u>														
Self												X		
School								X			X		X	X
Both	X	X	X	X	X	X	X		X	X				
<u>(Sub)Dimensions Assessed</u>	2	2	2	2	2	3	4	2	4	4	3	3	2	2
Academic	X					X	X		X	X	X	X		X
Leadership									x	x				x
Professional Development										x				
Teaching & Learning	x					x	x		x	x	x	x		x
Community	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Connectedness	x	x	x	x		x	x			x	x	x		x
Partnership							x	x		x	x		x	
Quality of Relationships	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Respect for Diversity							x	x	x	x		x	x	x
Institutional Environment							X		X	X				
Availability of Resources									x	x				
Environmental							x			x				
Structural Organization										x				
Safety		X	X	X	X	X	X	X	X	X	X	X	X	
Order & Discipline		x		x	x	x	x	x	x	x	x	x	x	
Physical		x	x	x	x	x		x		x			x	
Social/Emotional		x		x		x	x	x		x	x		x	
<u>Model</u>														
One-Factor			X	X	X	X						X		X
Multi-Factor	X						X	X		X				
Second-Order	X								X		X			
Bi-Factor		X											X	
<u>Number of Items</u>	14	23	5	11	7	9	42	24	21	81	15	7	26	7
<u>Number of (Sub)Factors*</u>	3	5	1	1	1	1	10	7	5	2	2	1	7	1

Table 1.9

Methodological Quality of Survey Validation Checklist – Individual Level

School Climate Measure	Author(s) (year)	Study	Structural Validity			Internal Consistency			Invariance		Construct Validity	
			1	2	3	4	5	6	7	8	9	10
Student Surveys												
Perceived School Experiences Scale (PSCS)	Anderson-Butcher et al., 2012	1	Good	Good	Doubtful	Inadequate	Good	Doubtful	Good	Doubtful	N/A	N/A
Delaware School Climate Survey-Student (DSCS-S)	Bear et al., 2011	1	Good	Good	Good	Good	Good	Doubtful	Good	Good	N/A*	N/A*
Add Health School Connectedness Scale (SCS)	Furlong et al., 2011	1	Good	Good	Doubtful	Good	Good	Doubtful	Good	Doubtful	Good	Doubtful
Georgia Elementary School Climate Survey	La Salle et al., 2016	1	Good	Good	Doubtful	Good	Adequate	Doubtful	N/A	N/A	Good	Doubtful†
Tripod School Climate Index	Phillips & Rowley, 2015	1	Adequate	Good	Good	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		2	Good	Good	Good	N/A	N/A	N/A	Good	Good	N/A	N/A
Georgia Brief School Climate Inventory (GaBSCI)	White et al., 2014	1	Good	Good	Doubtful	Good	Good	Doubtful	N/A	N/A	Good	Doubtful
School Climate Measure (SCM)	Zullig et al., 2015	1	Good	Good	Good	Good	Good	Good	Good	Good	N/A	N/A
Teacher/Staff Surveys												
Delaware School Climate Survey-Teacher/ Staff (DSCS-T/S)	Bear et al., 2014	1	Good	Good	Good	Good	Good	Doubtful	Good	Good	N/A*	N/A*
Revised School Level Environment Questionnaire (Revised SLEQ)	Johnson et al., 2007	1	Good	Good	Doubtful	Good	Good	Doubtful	Good	Doubtful	N/A	N/A
Teacher Questionnaire	Levitch et al., 2008	1	Good	Good	Doubtful	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brief, California School Climate Survey (Brief-CSCS)	You et al., 2014	1	Good	Good	Doubtful	Inadequate	Good	Doubtful	Good	Doubtful	N/A	N/A
Parent/Guardian Surveys												
Parent perceptions of school fit scale	Bahena et al., 2016	1	Good	Good	Doubtful‡	Good	Good	Doubtful‡	N/A	N/A	N/A	N/A
		2	Good	Good	Doubtful‡	Good	Good	Doubtful‡	N/A	N/A	Good	Doubtful‡
		3	Good	Good	Doubtful‡	N/A	N/A	N/A	Good	Doubtful‡	N/A	N/A
Delaware School Climate Survey-Home (DSCS-H)	Bear et al., 2015	1	Good	Good	Good	Good	Good	Doubtful	Good	Good	N/A*	N/A*
Parent perceptions of school climate scale	Schueler et al., 2014	1	Good	Good	Doubtful‡	Good	Good	Doubtful‡	Good	Doubtful‡	Good	Doubtful‡

*Conducted construct validity analyses at the school level.

†Utilized HLM to control for school effects but used total score from survey validated using raw response data.

‡May not apply, utilized national sample, but did not explain.

Table 1.10

Summary of Methodological Quality – Individual Level

School Climate Measure	Author(s) (year)	Methodological Quality				
		Structural Validity	Internal Consistency	Measurement Invariance	Construct Validity	Total
Student Survey						
Perceived School Experiences Scale (PSCS)	Anderson-Butcher et al., 2012	Doubtful	Inadequate	Doubtful	N/A	2
Delaware School Climate Survey-Student (DSCS-S)	Bear et al., 2011	Good	Doubtful	Good	N/A*	5
Add Health, School Connectedness Scale (SCS)	Furlong et al., 2011	Doubtful	Doubtful	Doubtful	Doubtful	4
Georgia Elementary School Climate Survey	La Salle et al., 2016	Doubtful	Doubtful	N/A	Doubtful†	3
Tripod School Climate Index	Phillips & Rowley, 2015	Good	N/A	Good	N/A	4
Georgia Brief School Climate Inventory (GaBSCI)	White et al., 2014	Doubtful	Doubtful	N/A	Doubtful	3
School Climate Measure (SCM)	Zullig et al., 2015	Good	Good	Good	N/A	6
Teacher/Staff						
Delaware School Climate Survey-Teacher/ Staff (DSCS-T/S)	Bear et al., 2014	Good	Doubtful	Good	N/A*	5
Revised version of the School Level Environment Questionnaire (Revised SLEQ)	Johnson et al., 2007	Doubtful	Doubtful	Doubtful	N/A	3
Teacher Questionnaire	Levitch et al., 2008	Doubtful	N/A	N/A	N/A	1
Brief, California School Climate Survey (Brief-CSCS)	You et al., 2014	Doubtful	Inadequate	Doubtful	N/A	2
Parent/Guardian						
Parent perceptions of school fit scale	Bahena et al., 2016	Doubtful‡	Doubtful‡	Doubtful‡	Doubtful‡	4
Delaware School Climate Survey-Home (DSCS-H)	Bear et al., 2015	Good	Doubtful	Good	N/A*	5
Parent perceptions of school climate scale	Schueler et al., 2014	Doubtful‡	Doubtful‡	Doubtful‡	Doubtful‡	4

*Conducted construct validity analyses at the school level.

†Utilized HLM to control for school effects but used total score calculated using raw data.

‡May not apply, utilized national sample, but did not explain.

Table 1.11

Methodological Quality of Survey Validation Checklist – School Level

School Climate Measure	Author(s) (year)	Study	Structural Validity			Internal Consistency			Invariance		Construct Validity	
			1	2	3	4	5	6	7	8	9	10
Student Surveys												
Perceived School Experiences Scale (PSCS)	Anderson-Butcher et al., 2012	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Delaware School Climate Survey-Student (DSCS-S)	Bear et al., 2011	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Good	Doubtful
Add Health School Connectedness Scale (SCS)	Furlong et al., 2011	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Georgia Elementary School Climate Survey	La Salle et al., 2016	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Tripod School Climate Index	Phillips & Rowley, 2015	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Georgia Brief School Climate Inventory (GaBSCI)	White et al., 2014	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
School Climate Measure (SCM)	Zullig et al., 2015	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Teacher/Staff Surveys												
Delaware School Climate Survey-Teacher/ Staff (DSCS-T/S)	Bear et al., 2014	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Good	Doubtful
Revised School Level Environment Questionnaire (Revised SLEQ)	Johnson et al., 2007	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Teacher Questionnaire	Levitch et al., 2008	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Brief, California School Climate Survey (Brief-CSCS)	You et al., 2014	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Parent/Guardian Surveys												
Parent perceptions of school fit scale	Bahena et al., 2016	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Delaware School Climate Survey-Home (DSCS-H)	Bear et al., 2015	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Good	Doubtful
Parent perceptions of school climate scale	Schueler et al., 2014	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

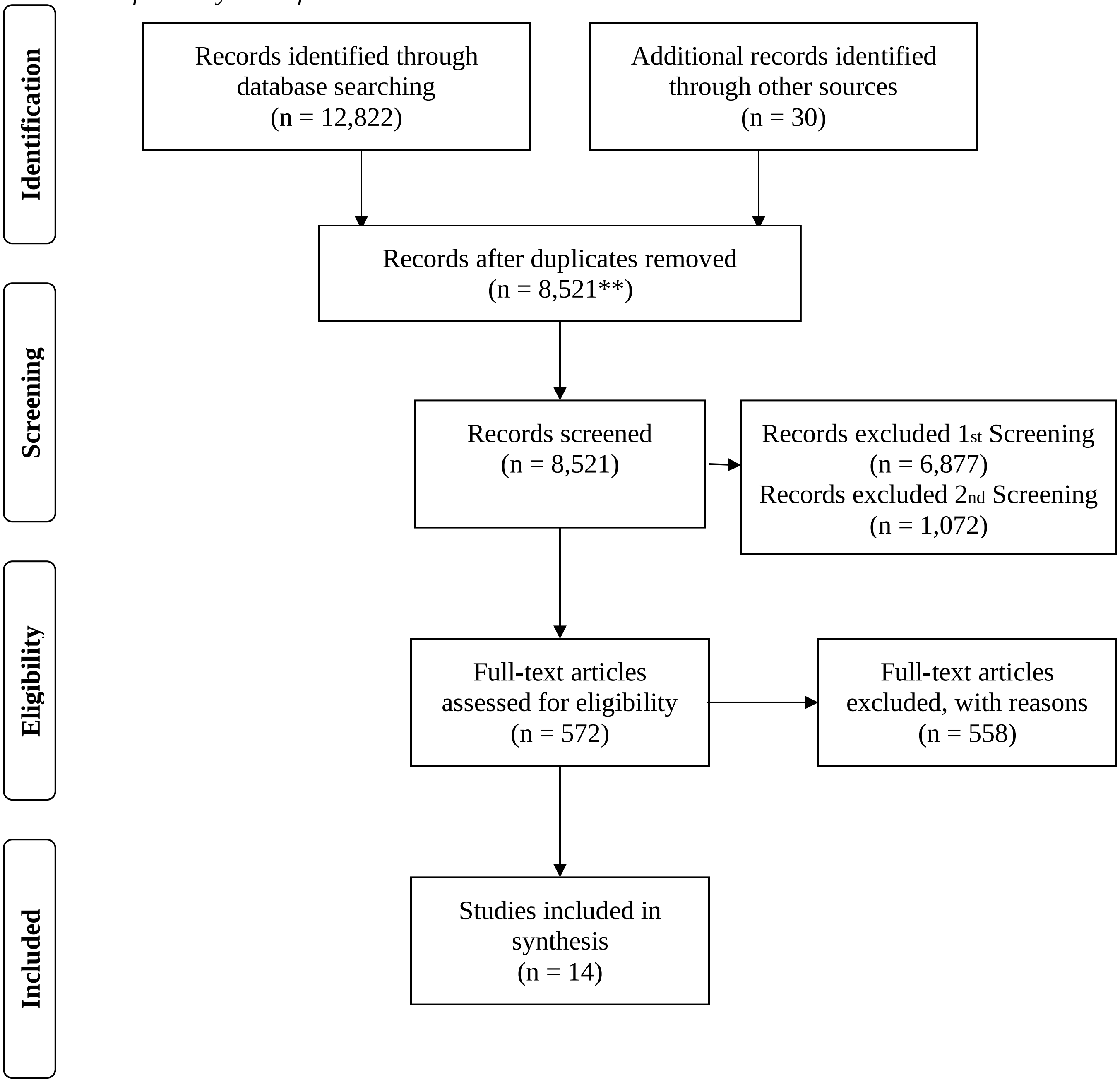
Table 1.12

Summary of Methodological Quality – School Level

School Climate Measure	Author(s) (year)	Methodological Quality				
		Structural Validity	Internal Consistency	Measurement Invariance	Construct Validity	Total
Student Survey						
Perceived School Experiences Scale (PSCS)	Anderson-Butcher et al., 2012	N/A	N/A	N/A	N/A	0
Delaware School Climate Survey-Student (DSCS-S)	Bear et al., 2011	N/A	N/A	N/A	Doubtful	1
Add Health, School Connectedness Scale (SCS)	Furlong et al., 2011	N/A	N/A	N/A	N/A	0
Georgia Elementary School Climate Survey	La Salle et al., 2016	N/A	N/A	N/A	N/A	0
Tripod School Climate Index	Phillips & Rowley, 2015	N/A	N/A	N/A	N/A	0
Georgia Brief School Climate Inventory (GaBSCI)	White et al., 2014	N/A	N/A	N/A	N/A	0
School Climate Measure (SCM)	Zullig et al., 2015	N/A	N/A	N/A	N/A	0
Teacher/Staff						
Delaware School Climate Survey-Teacher/ Staff (DSCS-T/S)	Bear et al., 2014	N/A	N/A	N/A	Doubtful	1
Revised version of the School Level Environment Questionnaire (Revised SLEQ)	Johnson et al., 2007	N/A	N/A	N/A	N/A	0
Teacher Questionnaire	Levitch et al., 2008	N/A	N/A	N/A	N/A	0
Brief, California School Climate Survey (Brief-CSCS)	You et al., 2014	N/A	N/A	N/A	N/A	0
Parent/Guardian						
Parent perceptions of school fit scale	Bahena et al., 2016	N/A	N/A	N/A	N/A	0
Delaware School Climate Survey-Home (DSCS-H)	Bear et al., 2015	N/A	N/A	N/A	Doubtful	1
Parent perceptions of school climate scale	Schueler et al., 2014	N/A	N/A	N/A	N/A	0

Figure 1.1

*Flow Chart for Study Identification and Selection**



*From: Moher, Liberati, Tetzlaff, & Altman, The PRISMA Group (2009).

**8,618 records were identified after electronically identifying and removing duplicates. An additional 127 duplicates were removed in the initial title screening process.

2 EXAMINING THE MULTILEVEL FACTOR STRUCTURE AND INVARIANCE OF THE GEORGIA SCHOOL CLIMATE SCALE

School climate has been conceptualized as "the heart and soul of the school. It is about that essence of a school that leads a child, a teacher, an administrator, a staff member to love the school and to look forward to being there each school day" (Freiberg & Stein, 1999, p.11). More concretely, school climate includes the school's "… norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures" (National School Climate Council, 2007, p. 4). It encapsulates nearly every aspect of the school environment and shapes the experiences and interactions of all school stakeholders, including students, teachers, parents, and staff. School climate has been recognized as an opportunity to foster student success due to its demonstrated links to desirable academic, social/emotional, and behavioral outcomes and its critical role in the school improvement process (Wang & Degol, 2016). The significance of school climate and the value of its study have been made clear both in educational literature (see, e.g., Thapa et al., 2013) and educational policy (see, e.g., Every Student Succeeds Act [ESSA], 2015).

School climate is a broad construct that encompasses multiple dimensions of school life. In a recent systematic review of hundreds of school climate research articles, Wang and Degol (2016), developed a conceptualization and categorization of school climate that identified the following domains: *Safety* (e.g. Social/Emotional, Discipline & Order, and Physical), *Community* (e.g. Partnership, Quality of Relationships, Connectedness, and Respect for Diversity), *Academic* (e.g., Leadership, Teaching & Learning, Professional Development), and *Institutional Environment* (e.g. Environmental, Structural Organization, and Availability of Resources). Reflecting school climate's importance is a vast body of research investigating its impact and relationships with outcomes. Within this body of research, particular attention has been paid to methods of measuring school climate. In their review of related literature, Wang and Degol (2016) found that roughly 15% of identified studies focused solely on developing and validating school climate measures. Given its complexity, the consideration paid to its measurement is appropriate. Still, the majority of these studies utilize conventional analyses to investigate school climate measures, such as single-level internal consistency estimates and factor analysis validity tests (Ramelow et al., 2015). Additional consideration using advanced analytic strategies is needed to accurately examine the psychometric properties of school climate measures and the validity of relationships determined using such measures (Konold et al., 2014; Marsh et al., 2009, 2012; Morin et al., 2014; Schweig, 2014; Wang & Degol, 2016).

Uses of School Climate Measures

As reflected in its inclusion in the "Every Student Succeeds Act" (ESSA: 2015), more and more states are reporting school climate indicators alongside more traditional academic outcomes within their accountability systems (e.g., the School Climate Star Rating index: Georgia Department of Education [GaDOE], 2015). Several statewide initiatives to measure school climate have been established in response (see: the *California School Climate, Health, and Learning Survey* (CAL-SCHLS) *System*: California Department of Education [CDE], 2017; the *Delaware School Surveys*: Bear et al., 2016; and the *Georgia School Climate Survey Suite*: GaDOE, 2016). Results from such surveys, aggregated to the school level, are often provided to schools for self-assessment and program development and/or evaluation purposes (Bear et al., 2016; GaDOE, 2015). For example, schools may use these results to identify areas in need of improvement, to develop targeted interventions in response to areas of concern, or to evaluate established programs, such as Positive Behavior Interventions and Supports (PBIS). In addition, aggregated school climate reports are often made available online for prospective parents or the general public to access (GaDOE, 2015). These reports may impact school reputations and/or influence decisions regarding where parents choose to live or send their children to school. Accordingly, the stakes attached to the accurate measurement of school climate are greater than ever.

Unfortunately, the complexity of school climate presents an array of challenges when attempting to accurately measure it. School climate is a multifaceted, multilevel construct experienced by multiple groups of school stakeholders within diverse school settings. It is overwhelmingly measured using survey data, from which several analytic issues arise, such as: the inclusion and/or exclusion of specific dimensions, the use of different reporters (e.g., parents, teachers and so forth), the clustered nature of survey data, and the equality of surveys for diverse populations (Bear et al., 2016; Konold et al., 2014; Konold & Cornell, 2015; Wang & Degol, 2016; Zabek et al., 2017). Some of these issues, particularly the inclusion of the multiple dimensions of school climate in surveys and the use of different reporters, are progressively reflected in literature and in practice. For example, comprehensive measures that examine several dimensions of school climate have become standard in statewide surveys (see, e.g., *California Healthy Students Survey*: Furlong et al., 2005; *Delaware School Survey – Student*: Bear et al., 2011*; Georgia School Climate Survey*: GaDOE, 2016). In addition, although student perceptions of school climate still make up the majority of the literature (Wang & Degol, 2016), surveys measuring the perceptions of adult stakeholders such as parents, teachers, and staff have been increasingly introduced (see, e.g., *California School Staff Survey* and *California School Parent Survey*: CDE, 2017; *Delaware School Climate Survey—Teacher/Staff*: Bear et al., 2014:

Delaware School Climate Survey—Home: Bear et al., 2015; *Georgia School Personnel Survey* and *Georgia Parent School Climate Survey*: GaDOE, 2015).

The aforementioned advancements regarding the inclusion of multiple dimensions of school climate and the use of different reporters are promising. However, significantly greater attention is required regarding the complex analyses recommended to address the clustered (multilevel) nature of school climate data and the equality (invariance) of school climate surveys for diverse student and school populations (Konold et al., 2014; Marsh et al., 2009, 2012; Morin et al., 2014; Schweig, 2014; Wang & Degol, 2016; Zabek et al., 2017). In addition, a more thorough theoretical foundation to guide school climate research is needed (Ramelow et al., 2015; Wang & Degol, 2016). Particularly, a conceptual model to justify decisions regarding the inclusion of particular levels (e.g. individual and school level) and groups (e.g. gender, racial/ethnic, and age groups) in school climate analyses is required. The present study will employ a bioecological framework (Bronfenbrenner & Morris, 2006) to investigate the multilevel nature and invariance of a school climate survey using multilevel confirmatory factor analysis (MCFA) and multilevel multiple indicators, multiple causes (MIMIC) modeling procedures.

Theoretical Foundation

Researchers have noted that theory-grounded measurement development is often missing in school climate literature (Konold et al., 2014; Ramelow et al., 2015; Wang & Degol, 2016). A clear link between theory, methods of measurement, and statistical analyses is needed. Bronfenbrenner's *bioecological model of human development* (Bronfenbrenner & Morris, 2006) provides a framework within which to understand the relationships among school climate domains, explain its influence on outcomes, and formulate strategies for school improvement.

Briefly, this theory posits that developmental outcomes are influenced by the combined interactions among personal characteristics, proximal processes, contexts, and time. It aligns with the conceptualization of school climate as a complex phenomenon that reflects the interactions among multiple stakeholders and dimensions of the school environment. In addition, careful adherence to Bronfenbrenner's theory in its "mature" form (e.g. see, Bronfenbrenner & Morris, 2006) can help clarify the statistical techniques required to ensure that decisions regarding school climate's measurement and analysis are grounded in theory.

The bioecological model is comprised of four defining properties – (1) Process, (2) Person, (3) Context, and (4) Time – referred to as the PPCT Model when applied to an operational research design (Bronfenbrenner & Morris, 2006). *Process* is the core of the model. Specifically, *proximal processes* are posited as the primary engines of development and encompass the interactions between an individual and her immediate environment (i.e. people, objects, and symbols) and vary systematically as a joint function of person, context, and time characteristics. *Person* refers to the personal characteristics of an individual (e.g. gender, cognitive ability, and temperament). *Context* refers to the characteristics of the four interrelated systems that were described in Bronfenbrenner's early work (i.e., microsystem, mesosystem, exosystem, and macrosystem: for a detailed description, see, Bronfenbrenner, 1977). Lastly, *Time* refers to the stability and periodicity of proximal processes, as well as to changes in expectations and societal norms over time. While the PPCT model emphasizes the assessment of time to better understand the mechanisms and conditions of development, it will not be addressed in the current study.

Critical to school climate research is the PPCT model's distinction between *interactions* with a setting (proximal processes) and *characteristics* of it (microsystems). While the safety, community, institutional environment, and academic dimensions of school climate have always been clearly contained within Bronfenbrenner's microsystem, the PPCT model further clarifies the ways in which those dimensions influence outcomes. Thus, it provides a conceptual framework to guide the interpretation and modeling of individual- and school-level climate (see Figure 2.1). For example, individual-level climate variables that directly assess an individual's interactions with the school may represent proximal processes. At the same time, school-level climate variables that reflect the shared characteristics of the school may be conceived as a characteristic of the school microsystem. The interactions between these variables and their influence on outcomes may vary; therefore, it is important they are modeled and investigated separately. To do so, it is critical to account for the nested nature of the data and validate the factor structure at the level of interest. The present study will employ multilevel confirmatory factor analysis (MCFA) techniques to address these issues.

A core component of the PPCT model is the investigation of the interrelations among elements. In addition to assessing how proximal processes may vary systematically as a function of person, context, and time characteristics, it also considers the relationships among and interaction effects of any combination of its four defining principles. Expected relationships based on theory or previous research should be hypothesized and examined. Thus, the PPCT model provides a conceptual framework for investigating the influences of process, person, context, and time phenomena on school climate and its relationships. For example, based on previous research, one might expect individual-level climate to vary systematically as a function of culture based on research that shows a relationship between perceptions of climate and racial/ethnic group membership (Battistich et al., 1995; Griffith, 2000; Konold et al., 2017; Kuperminc et al., 1997). Similarly, one may hypothesize school-level climate to vary as a function of other characteristics of the school microsystem based on research that shows a link between school climate and school racial composition (e.g., see, Konold, et al., 2017) and school socioeconomic status (SES) (e.g., see, Battistich et al., 1995; Vieno et al., 2005). However, to ensure group differences are not due to measurement error, it is important to first ensure the equality of the individual- and school-level survey across the diverse populations being assessed. The current study will utilize multilevel multiple indicator multiple cause (MIMIC) modeling techniques to investigate the invariance of a school climate measure across these diverse populations and subsequently analyze group differences.

Issues of Clustered Data

While school climate is often investigated at the individual level, theoretically and in practice, school climate is most often conceptualized at the school level. Theoretically, it is defined as the overarching dimensions that encapsulate nearly every aspect of the school environment and shape the experiences of all school stakeholders (Cohen et al., 2009; Thapa et al., 2013; Wang & Degol, 2016). In practice, it is reported at the school level as an indicator of school quality and positive functioning. Still, school climate is primarily measured using surveys of the perceptions of individuals, such as students or other school stakeholders. The resulting data are inherently hierarchical – consisting of individuals clustered within schools – and confusion arises when evaluating the meaning and psychometric properties of such surveys (Konold et al., 2014; Marsh et al., 2009, 2012; Schweig, 2014; Wang & Degol, 2016). This confusion stems from what Bliese (2000) refers to as “fuzzy” variables. Fuzzy variables are simultaneously influenced by individual- and environmental-factors, and therefore, cannot be considered purely level-1 or level-2 constructs. Regarding school climate surveys, student responses represent fuzzy variables because they reflect both individual differences in perceptions and common school factors. Thus, responses fail to meet the assumption of data independence.

Often in research and practice, a conventional method is followed to investigate school climate and validate its measures that does not account for the inherent dependency of the data. Individuals' observed responses on school climate surveys are analyzed to determine the homogeneity and intercorrelations of items and factors (Konold et al., 2014). Factor scores are subsequently used to examine group differences (e.g. gender differences) in perceptions of school climate and to determine the associations between school climate and other constructs at the individual level. Typically, individual responses are then aggregated to the school level and interpreted as indicators of a school's climate. These scores are used to assess school functioning, to explore how climate at the school-level relates to outcomes, and to compare schools in terms of their climate. This conventional method of school climate measurement reflects two critical problems outlined in Morin et al. (2014): (1) the failure to appropriately identify and differentiate the desired level(s) of analysis, and (2) the failure to control for measurement and sampling error when conducting analyses.

Due to the nested nature of school climate data, analyses can be conducted at the individual and/or school level. However, it is critical to recognize that school-level and individual-level school climate may reflect two separate and distinct constructs. Constructs often have different meanings at different levels of analysis (Bliese, 2000). Further complicating interpretation, the meaning at each level may depend on whether the constructs are composed of items that reference the self or the group (Chan, 1998). *Self-referent* items (e.g. "I treat other students fairly") are designed to reflect individual-level constructs that then may be used to infer school-level constructs. *Group-referent* items (e.g. "Students treat one another fairly") are designed to reflect group-level constructs but may reflect individual differences. The meaning of a construct at the individual- and school-level may differ depending on whether it is based on self- or group-referent items (for more information, see, e.g., Bliese, 2000, Chan, 1998, Lüdtke et al., 2008; Marsh et al., 2009; Morin et al., 2014). In either case, interpreting individual-level school climate as an accurate reflection of school-level climate, as outlined in the previous paragraph, is an example of the *ecological fallacy* (Cronbach, 1976; Robinson, 1950).

The ecological fallacy is a common phenomenon in educational research (Morin et al., 2014). It mistakenly assumes that observations and findings at one level of analysis generalize to another. Marsh et al. (2009) illustrates the ecological fallacy using the *big-fish-little-pond effect*, a classic contextual effect in which the direction of association between academic achievement and academic self-concept is opposite across levels. Consider Morin et al.'s (2014) explanation:

> Research shows that achievement at the individual-student level has a positive effect on academic self-concept ("The brighter I am, the better my academic self-concept"), but school- or classroom-average achievement has a negative effect on academic self-concept ("The brighter my classmates, the lower my academic self-concept"). (p. 145)

This illustration demonstrates how using conventional strategies to explore school climate may result in the ecological fallacy. The school-level climate phenomenon may be completely unrelated to the individual-level climate phenomenon (Longford & Muthén, 1992). For example, Marsh et al. (2012) found that classroom-level climate has a significant effect on student achievement: "Students achieve less in classrooms perceived to be chaotic" (p. 116). However, student-level climate had no meaningful relationship with achievement at the individual level. In relation to school climate surveys, relationships among items and factors – including the internal consistency estimates, the number of factors extracted, the magnitudes and patterns of factor loadings, and the overall factor structures – may be very different at the individual level and the school level.

Failure to account for the hierarchical structure when evaluating factor analytic and structural models of school climate can result in incorrect conclusions and interpretations (Konold et al., 2014). Not only do the individual-level school climate construct and the school-level climate construct need to be explored separately, but also, the measurement and sampling error associated with nested data needs to be controlled for when conducting such analyses (Morin et al., 2014; Muthén, 1991, 1994; Muthén & Asparouhov, 2011). In the previously described conventional method of investigation, student-level school climate was examined using observed variables (individual responses to survey items) and school-level climate was examined using aggregated scores. The use of observed variables in this context not only fails to meet the assumptions of independence, but also, the use of aggregated scores additionally fails to account for variations in sample size across schools.

Observed variables are comprised of both individual and school effects – scores for individuals within the same school are not independent (Muthén, 1994). At the individual level, using observed scores fails to account for the shared influence of school factors. At the school-level, using aggregated scores fails to control for the within-variability due to personal factors. Meaningful variability across individuals within schools may be due to differences in individual personalities and personal perceptions, and not a reflection of school effects. Variation across individuals between schools may be affected by differences in school environments and community characteristics, and not due solely to personal factors. At the same time, similarities among individuals within schools may be due to shared culture and experiences (Konold et al., 2014). Thus, to control for measurement error, the unique individual- and school-level effects need to be teased out before investigating school climate at each level (Konold et al., 2014; Muthén, 1994; Muthén & Asparouhov, 2011). In addition, the variability in number of individuals represented within different schools needs to be accounted for to control for sampling error (Morin et al., 2014).

Advancements in statistical modeling, particularly multilevel factor analytic and structural modeling techniques, offer strategies that can control for the measurement and sampling error associated with clustered data (Muthén, 1991, 1994; Muthén & Asparouhov, 2011). Regarding measurement error, conventional factor analytic strategies are conducted on a single covariance matrix created from observed or aggregated variables. Consequently, conventional methods confound the effects unique to each level within one matrix. In contrast, multilevel factor analytic techniques disaggregate each indicator's observed score into independent within-school and between-school components, creating two distinct covariance matrices. The *between matrix* represents the across-school variation that is independent of the within-school variance. The *within matrix* represents the within-school, individual-level variation that is independent of the school effect (Muthén, 1991, 1994; Muthén & Asparouhov, 2011). Separate models at each level of analysis can be estimated separately or simultaneously using the within and between matrices, thus, controlling for the measurement error inherent to nested data.

The ability to estimate separate models at each level also responds to the aforementioned ecological fallacy concern. First, it allows for the factor structure and relationships among indicators to be different for the individual- and school-level climate constructs. Next, it permits investigations of school climate's relationships with outcomes to be conducted separately on the individual- and school-level constructs without confounding the effects of each level. In addition, these multilevel methods permit the scores for various individuals within different schools to serve as multiple indicators of the latent school-level climate factors. By taking into consideration the number of individuals represented within each school, multilevel factor analytic and structural modeling techniques can account for the sampling error associated with clustered data in addition to controlling for the measurement error (Morin et al., 2014).

Issues of Cross-Level Invariance

As explained above, interpreting individual-level school climate as an accurate reflection of school-level climate assumes the equality of the school climate constructs across levels – an ecological fallacy (Morin et al., 2014). Statistically, this assumption is referred to as *cross-level invariance* (Dedrick & Greenbaum, 2011; Schweig, 2014; Zyphur et al., 2008). When data are clustered, cross-level invariance indicates parameter invariance across the levels of analysis that exist in the data – a constraint that is rarely met (Zyphur et al., 2008). To investigate the assumption of cross-level invariance, the multilevel methods described in the previous section can be extended to test the equality of constructs and factor structures across levels (Jak, 2013; Jak et al., 2013). Unfortunately, cross-level invariance is regularly assumed without first being checked, leading to an array of problems (Schweig, 2014; Zyphur et al., 2008). Failing to separately analyze the within- and between-covariance matrices assumes that the reliability estimates and factor structures are invariant across the individual- and school-level. Zyphur et al. (2008) describe the problems with such an approach as twofold.

First, it fails to discover similarities and differences across levels of analysis in the functioning of variables (Zyphur et al., 2008). Such differences could have important theoretical implications within school climate research. Cronbach (1976) explains that examining whether individual students who describe a class as apathetic also describe the class as difficult is different than examining if, when students collectively describe a class as apathetic, do they also describe it as difficult. “The former refers to the phenomenology of the student compared to other students rating the same events. The latter refers to behavioral differences between classes” (Cronbach, 1976, p. 9.19). Sirotnik (1980) posits that understanding the various variables (e.g., gender, race/ethnicity, length of employment) that account for differences in teacher influence is a worthy endeavor, and so too is understanding why some schools have greater teacher influence than others. When investigating the latter, it is important that the magnitude of teacher influence for the school be an attribute of the school and not of the teachers within the school. By unjustifiably imposing cross-level invariance, researchers miss the opportunity to understand constructs across levels.

Second, researchers use the results of single-level analyses to justify school climate items and factors, which are then often used to investigate its relationships at multiple levels of analysis (Zyphur et al., 2008). For example, the school climate factors validated using single-level analyses are often then aggregated to the school level to investigate their relationships to outcomes (e.g., see, Bear et al., 2011). These analyses make assumptions about the equality of the school climate factor structure at each level. In other words, they assume that school climate is measured in the same way, equally well, at the individual- and school-levels. Schweig (2014) found that the factor structure of a classroom climate survey differed significantly across levels of analysis, and that single-level factor analyses conducted on the observed correlation matrix distorted the school-level climate factor structure. Thus, the assumption of equality in factor structures across levels may not be valid, and the meanings of relationships discovered at each level are difficult to interpret.

These aforementioned cross-level assumptions are commonplace in school climate research and practice, and in educational policy more broadly (Schweig, 2014). With a few notable exceptions (e.g. see, Konold et al., 2014; Konold & Cornell, 2015; Schweig, 2014), the majority of research findings related to school climate assume cross-level invariance without first investigating it. As the inclusion of school-level climate as an indicator of school quality continues to increase within educational accountability systems, the trend of assuming cross-level invariance is troubling. Incorrectly assuming cross-level invariance has notable consequences. Depending on which level analyses are conducted, different dimensions of school climate may be defined. These dimensions often have a significant impact on policy and practice. By using the incorrect level of analysis to determine dimensions, qualities of school climate may be incorrectly targeted in policy and practice, or may be missed altogether (Schweig, 2014).

Given that cross-level invariance is often assumed in school climate research and policy practices, there is a need for greater exploration of the accuracy and impact of such assumptions. Multilevel factor analytic techniques provide a framework to evaluate whether school climate constructs have consistent meanings and measurement structures at the individual- and school-levels (Konold et al., 2014; Muthén, 1994; Muthén & Asparouhov, 2011; Schweig, 2014). By disaggregating the variation unique to each level as described in the previous section, separate factor structures of the individual- and school-level climate constructs can be estimated and examined for configural and metric invariance across levels (Jak, 2013; Jak et al., 2013; Meredith, 1993; Schweig, 2014; Zyphur et al., 2008).

First, *cross-level configural invariance* is investigated to determine the equality of the number of factors and pattern of factor loadings across levels. The factor structure at each level can be explored in two ways. These analyses can be conducted separately using the covariance matrices for each level, or simultaneously using MCFA (Jak, 2013; Muthén & Asparouhov, 2011; Schweig, 2014; Zyphur et al., 2008). Within a multilevel model, to better understand the factor structure at a certain level, an unrestricted model can be specified at the other level (Muthén & Asparouhov, 2011). While the magnitude of factor loadings is also considered when investigating configural invariance, larger between-level factor loadings are expected due to the smaller sample size (Zyphur et al., 2008). Support for cross-level configural invariance suggests that the latent school climate factors are conceptually comparable across levels.

If support for configural invariance is found, *cross-level metric invariance* is investigated to determine the equality of factor loadings across levels (Bottoni, 2016; Jak, 2013; Mehta & Neale, 2005). An initial MCFA is specified in which all factor loadings are freely estimated. Then, a second MCFA is specified in which the factor loadings are constrained to equality across levels. If the constraint of equal factor loadings does not significantly worsen model fit, support for metric invariance is found. Conceptually, metric invariance suggests the school climate factors are being measured in the same way at each level (Selig et al., 2008). In other words, the latent climate factors can be interpreted similarly (Jak, 2013). In addition, the latent school climate variances can be compared across levels, because constraining the loadings equates the scale across levels (Mehta & Neale, 2005; Selig et al., 2008).

If support for configural and metric invariance is found, it suggests the school climate construct and factor structure possess cross-level measurement invariance. However, this does not imply that the latent factors at each level can be validly compared across diverse individual and school populations. To assess the equality of the factor structure across schools, an additional set of constraints can be specified to determine the presence of *cluster bias* (Jak et al., 2013). To test for cluster bias, Jak (2013) recommends the residual variances at the between level are constrained to zero. If this constraint significantly worsens model fit compared to the metric model, modification indices are examined to determine the source of bias. One by one, the between level residual variances of suggested indicators are freed until the model fits the data well.

The presence of cluster bias indicates that the survey does not measure school climate equally across schools, particularly in the items that were freely estimated and their corresponding latent factors (Jak, 2013). In other words, two students who possess similar perceptions of school climate but attend two different schools may differ significantly with respect to their expected response on those items. Cluster bias suggests that one or more school-level characteristics are violating measurement invariance at the between level. For example, students at schools with more diverse racial/ethnic compositions may respond differently to one or more items on the survey when compared to students from schools with more homogeneous racial/ethnic compositions. When cluster bias is detected in a survey, school-level factors (e.g. student-teacher ratio, socioeconomic status composition, location) can be explored to determine if they account for the cluster bias (Jak, 2013). Issues related to measurement invariance with respect to individual- and school-level factors within multilevel models will be discussed in more detail below.

Issues of Student- and School-Level Invariance

The previous sections summarize the issues associated with clustered data and strategies to address those issues using multilevel factor analysis. Once the school climate factor structures are established at each level, it is important to determine their equality across diverse student and school populations. That is, it is important to determine whether the measurement structures possess *invariance* across groups. Invariance with respect to group membership has gained attention in educational research and policy. Invariance procedures are particularly useful to provide evidence that statistically the same construct is being measured, in the same way (i.e., without bias), for different groups (Chen, 2008). For example, it ensures that, for two students with identical perceptions of school climate, the probability of an observed response to a survey item is equal regardless of individual factors (e.g. regardless of their gender or race) or school factors (e.g. regardless of their school's population or student-teacher ratio) (E. Kim et al., 2012). Prior research has tested mean differences or moderation of latent factor means without evaluating measurement invariance—such that theoretical interpretations are conflated with measurement artifacts (Chen, 2008). If measurement invariance is not present, group similarities or differences on constructs may be the result of measurement error. As a result, comparisons and substantive implications of concluding group similarities or differences may not be valid.

Historically, school climate literature often concluded group similarities and differences in perceptions of school climate without first establishing measurement invariance in reference to those groups. More recently, testing invariance has become increasingly common before comparing latent group means in school climate research (see, e.g., Bear et al., 2011, 2015; Konold et al., 2014; Zabek et al., 2016). However, the issues associated with testing invariance of multilevel data, like school climate data, are just beginning to be explored in applied research (see, e.g., Jak, 2013; E. Kim et al., 2012, 2015). Thus, multilevel approaches to testing invariance have yet to be reflected in school climate literature. Ignoring the multilevel structure of data when conducting invariance analyses entails an array of assumptions similar to those described in the above sections. Particularly, it assumes cross-level invariance of the factor structure (E. Kim et al, 2012). In addition, by failing to separate the within and between components of the data, results confound the effects at each level. E. Kim et al. (2012) found that ignoring the multilevel structure inflates Type I error rates; that is, "invariant models are overly rejected and concluded to be noninvariant when the dependency of the data is not taken into account" (p.

265). Subsequently, the accuracy of invariance findings is unclear when tests were conducted on observed or aggregated variables. Thus, invariance with regard to individual- and school-level groups needs to be demonstrated using multilevel modeling techniques to ensure psychometrically valid comparisons of group differences are warranted.

Recent studies have evaluated the use of several strategies for testing invariance with respect to individual- and school-level factors (see, e.g., Jak, 2013; E. Kim et al., 2012, 2015). These approaches generally take two forms: (1) multilevel factor analytic strategies such as *multiple-group confirmatory factor analysis* (MGFA: Sörbom, 1974), and (2) multilevel structural equation modeling approaches such as *multiple indicators multiple causes analysis* (MIMIC: Muthén, 1989). The primary difference between each is MGFA tests for invariance, and MIMIC explores where invariance may be violated (E. Kim et al., 2012). Selig et al. (2008) compared these two methods and suggested MGFA when testing invariance in reference to a small number of groups and MIMIC modeling analyses when the number of groups becomes large. This is because MGFA assumes a finite number of groups; while MIMIC models can test invariance of continuous grouping variables (E. Kim et al., 2012).

Additional challenges arise when the groups of interest occur at the individual level (e.g. individual gender, race/ethnicity, and age groups). For example, multilevel MGFA techniques are sometimes used to investigate invariance with regard to individual-level groups (e.g., see, Lee et al., 2017). However, "constructing multigroup multilevel [MGFA] models for within-level groups is not feasible when the group indicator (e.g., females and males within schools) is crossed across higher level clusters" (E. Kim et al., 2015, p. 603-604). E. Kim et al. (2015) investigated alternative approaches for testing within-group invariance and found that both multilevel factor mixture model for known classes (ML FMM) and multilevel MIMIC (ML

MIMIC) procedures performed well. For consistency, Jak (2013) recommends a stepwise approach that uses MIMIC modeling to test individual- and school-level invariance, “so that the final model comprises all bias and substantive findings at both levels” (p. 91).

School climate research would benefit from investigations of the invariance of surveys using multilevel methods. Presently, school climate invariance analyses examine almost exclusively the invariance of measures at the individual level (see, e.g., Bear et al., 2011, 2015; Konold et al., 2014; Zabek et al., 2017). For example, Bear et al. (2011, 2015) centered their individual item responses around the school mean and conducted single-level invariance tests of their school climate factor structure across gender, grade level (e.g. elementary, middle, and high), and racial/ethnic groups. Taking a different approach, Konold et al. (2014) used multilevel factor analysis techniques to validate their measure of authoritative school climate, but then conducted structural invariance tests across gender groups using a single-level design. As explained above, this analysis assumes the cross-level invariance of the factor structure and may confound individual- and school-level effects. In addition, school climate research has thus far ignored the invariance of the school-level factor structures across diverse school populations. Invariance analyses with multilevel models provides a framework for exploring both individual- and school-level groups, the latter of which requires significantly more attention in school climate research. Due to the clustering of students within schools, noninvariance may be incorrectly concluded at the individual level instead of the school level if the dependency of the data is not considered. For example, if noninvariance in relation to student race/ethnicity is found, this may actually be an effect of school-level factors associated with schools of varying racial compositions.

Exploration of the invariance of school climate measures using multilevel models is especially important given the rise in school climate as an indicator for accountability purposes. States are required to disaggregate accountability information (e.g., school characteristic indicators such as school climate) by student subgroups (ESSA, 2015). Accurate accountability inferences rely on measurement tools working equally well across groups (e.g. gender, racial/ethnic, and grade level groups). Multilevel models are required to ensure this. In addition, if schools are assessed and compared in terms of their school climate, it is critical that the measures used not only work equally well across diverse student populations, but also across diverse school communities.

Present Study

The above sections reinforce calls for significantly greater exploration of the psychometric properties of school climate surveys in relation to the multilevel nature of the data and subsequent measurement invariance for diverse student and school populations (Konold et al., 2014; Marsh et al., 2009, 2012; Morin et al., 2014; Schweig, 2014; Wang & Degol, 2016; Zabek et al., 2017). The present study will employ multilevel factor analytic and structural modeling procedures (e.g. MCFA and ML MIMIC procedures) to respond to these calls. It will examine the multilevel structure, cross-level invariance, and individual- and school-level measurement invariance of the Georgia School Climate Survey (GSCS: GaDOE, 2016). The GSCS is a comprehensive measure of school climate for 6th through 12th grade students, administered annually to hundreds of thousands of public-school students throughout the state of Georgia. The GaDOE aggregates observed results from this measure to assess schools in terms of their climate and makes these school-level reports available online. Investigations of its multilevel structure and invariance are especially important to determine the validity of school-level reports.

Fist, MCFA procedures will be utilized to explore the multilevel factor structure of the GSCS. Two proposed models of the GSCS will be examined: the existing model (see, GaDOE, 2016) and an alternative model that sought to address some of the previously described issues of school climate measurement. For example, it aligns with a conceptual model of *school* climate, and thus excludes items from the existing GSCS model that are not specific to school (e.g., "*Honesty is an important trait to me*"). It also uses the categorization of school climate identified in Wang and Degol's (2016) systematic review to justify the inclusion and/or exclusion of specific school climate dimensions. The factor structures of the existing and alternative models will be examined to determine which model fits the data best and whether changes to the model are warranted. Then, the best fitting model will be investigated to explore its validity at each level, cross-level invariance, and equality across schools.

Second, multilevel MIMIC modeling procedures will be employed to explore the measurement invariance of the GSCS with respect to individual- and school-level factors. Specifically, the invariance across racial/ethnic groups at the individual level and racial/ethnic and SES composition at the school level will be investigated. While a vast array of individual- and school-level factors could be examined, these were chosen for several reasons. First, prior school climate research has frequently found that students identifying as racial/ethnic minorities perceive school climate less favorably than other students (see, e.g., Koth et al., 2008; Kuperminc et al., 1997; La Salle et al., 2016; White et al., 2014). Second, schools in the United States are becoming increasingly racially and ethnically diverse (U.S. Census Bureau, 2011, 2012). Third, despite the rising racial and ethnic diversity of the school age population, racial segregation is increasing in the nation’s schools (Frankenberg et al., 2003; Orfield & Frankenberg, 2014). Finally, children who belong to racial/ethnic minority groups are significantly more likely to live below the poverty line than other children (National Center for Education Statistics, 2013), and schools that serve high proportions of students living in poverty and students from racial/ethnic minority groups tend to have significantly less resources than other schools (U.S. Department of Education Office for Civil Rights, 2012).

Together, these circumstances highlight the need for multilevel investigations of the invariance of school climate measures in relation to student racial/ethnic groups and school racial/ethnic and SES composition. First, it is important to further explore whether the relationship between student racial/ethnic group membership and perceptions of school climate persist when measurement error and school-effects are taken into account. Due to the disproportionate clustering of racial/ethnic groups within schools, racial/ethnic differences in perceptions of school climate may be incorrectly concluded for an array of reasons. For example, it may be due to measurement noninvariance at the individual level. It may also be due to confounding school-effects specific to schools that predominantly serve racial/ethnic minority populations, or to measurement noninvariance with respect to corresponding school-level factors such as racial/ethnic and SES composition. After accounting for the multilevel nature of the data and measurement invariance across racial/ethnic groups, the present study seeks to examine whether mean differences observed in prior research on school climate are replicated within our sample.

In addition, the aforementioned disproportionate clustering of students within schools emphasizes the need to determine measurement invariance across school-level factors. The use of school climate as an indicator of school quality within accountability systems suggests schools will be compared in terms of climate. Thus, it is critical that measurement invariance with respect to school-level factors be established. The dimensions of school climate assessed within surveys may not be equally measured across schools of varying racial/ethnic and SES compositions. For example, “respect for diversity” is a subdimension of school climate (Wang & Degol, 2016), assessed within the Cultural Acceptance subscale of the GSCS. It is reasonable to hypothesize that respect for diversity is conceptually different in schools that are highly diverse versus schools that are relatively homogeneous. Similarly, “availability of resources” is a subdimension of school climate (Wang & Degol, 2016), assessed within the Physical Environment subscale of the GSCS. Again, the equality of this construct may differ between a school that predominantly serves students living below the poverty line and a school comprised of primarily upper-middle class students. Thus, the present study seeks to explore the measurement invariance of the GSCS with respect to these school-level factors, and subsequently determine the relationships between school climate and racial/ethnic and SES composition at the school level.

In summary, the purpose of the current study is to explore the multilevel factor structure of the GSCS and its invariance with respect to student (i.e. race/ethnicity) and school (i.e. racial/ethnic and SES composition) level variables. The goal of this investigation is to determine the psychometric properties of the GSCS in relation to these areas and determine whether changes to the scale may be appropriate. Accordingly, the following research questions will be addressed.

Research Question 1

1a. What GSCS factor structure best fits the survey data?

1b. Does the identified GSCS factor structure fit the within- and between-level data?

1c. Does the factor structure of the GSCS demonstrate cross-level invariance?

1d. Is the factor structure of the GSCS invariant across schools, or is cluster bias detected?

Research Question 2

2a. Does the multilevel factor structure of the GSCS demonstrate invariance with respect to student-level variables (i.e. across racial/ethnic groups)?

2b. Does the multilevel factor structure of the GSCS demonstrate invariance with respect to school-level variables (i.e. across schools of various racial/ethnic compositions and SES compositions)?

Research Question 3

3a. After accounting for any measurement error identified in Research Questions 1 and 2, what is the relationship of school climate and its subdimensions, as measured by the GSCS, with student- and school-level variables (i.e. student race/ethnicity, and school racial/ethnic composition and SES composition)?

Method

Design and Procedures

The GaDOE administers the Georgia Student Health Survey (GSHS) 2.0 annually to 6th through 12th grade public school students throughout the state of Georgia. The GSHS 2.0 is an anonymous, statewide, self-report instrument electronically administered online. It includes items measuring student perceptions of multiple social and emotional domains and contains the Georgia School Climate Survey (GSCS). All Georgia public schools are encouraged to participate in the GSHS 2.0, and at least 75% of students in each grade level must participate in the GSHS 2.0 for results to be included in the GaDOE's School Climate Star Rating.

The current study utilized data from the 2015-2016 school year iteration of the GSHS 2.0. School-level racial/ethnic and SES composition data were gleaned from the State of Georgia's Governor's Office of Student Achievement (GOSA: Georgia's GOSA, 2016a, 2016b). To control for school-level effects associated with school type (e.g. K-8, middle, high: see, e.g., H. Kim et al., 2014), only data from "traditional" middle schools (i.e. those serving exclusively 6th - 8th grade students) were included in analyses. According to GaDOE protocol, passive consent procedures were used, and data were collected anonymously and received in de-identified form from the GaDOE. The university institutional review board approved all study procedures.

Students that positively endorsed an item created to detect random response patterns were removed before data were received from the GaDOE. Data were removed from an additional 13,239 participants (4.8%) who were detected as providing inconsistent responses (i.e., provided conflicting answers on items asking whether they had seriously considered harming themselves in the past year [items 79 and 80] and/or on items asking whether they had been bullied/teased by other students [items 39 and 40]). Data from another 1,388 students (0.5%) were removed due to missing grade data. Finally, one school had less than 10 participants (n = 9), and the data from these participants were removed. The resulting sample of 259,778 students represented 427 middle schools from 147 school districts throughout Georgia.

Participants

The GSHS sample was comprised of 50.8% females (n = 132,053) and 49.2% males (n = 127,725). Participants represented the following racial/ethnic backgrounds, as identified on the GSHS: 42.7% White/Caucasian (n = 110,940), 33.6% Black/African American (n = 87,247), 13.7% Hispanic/Latino (n = 35,686), 4.7% Asian/Pacific Islander (n = 12,326), and 5.2% Other (n = 13,579). The grade-level distribution was: 33.2% sixth grade (n = 86,220), 33.5% seventh grade (*n* = 86,917), and 33.4% eighth grade (*n* = 86,641). The final sample was representative of statewide GaDOE demographic data from the same year (41% White, 37% Black, 15% Hispanic/Latino, 4% Asian, and 3% Multiracial).

There was an average of 608 participants from each of the 427 schools (range = 68 -1806, *SD* = 278, *Mdn* = 575). Schools were, on average: 41.6% White (range = 0% - 98%, *SD* = 28.4), 39.2% Black (range = 0% - 99%, *SD* = 29.6), 12.5% Hispanic (range = 0% - 88%, *SD* = 13.3), 3.3% Multiracial (range = 0% - 10%, *SD* = 1.7), 3.2% Asian (range = 0% - 54%, *SD* = 5.7), and 0.6% American Indian/Alaska Native (range = 0% - 2%, *SD* = 0.24). An average of 34.9% (range = 1% - 85%, *SD* = 18.5) students at each school directly qualify for free-or-reduced lunch.

Measures and Variables

Georgia School Climate Survey

There are 52 items that inquire about students' perceptions of school climate within *Section A: School Climate* (items 1-45 and 87-93) of the GSHS 2.0, 2015-2016 iteration. Using this pool of items, two possible factor structures were evaluated in the present study.

Georgia School Climate Survey – Original. The original *Georgia School Climate Survey* (GSCS) factor structure was developed using a series of EFA and CFA procedures (GaDOE, 2016). It is a measure of school climate that consists of 36 Likert-scale items with response options ranging from '1' (Strongly Disagree) to '4' (Strongly Agree; see Table 1). Previous factor analyses conducted at the individual level using observed responses from the Fall 2014 iteration of the GSCS found support for a second-order model of school climate (GaDOE, 2016). This model revealed an overall school climate factor representing the relationships between eight subfactors: school connectedness (5 items), peer social support (3 items), adult social support (4 items), cultural acceptance (5 items), character (6 items), physical environment

(4 items), safety (4 items), and order and discipline (5 items) (see Table 1). Model fit statistics provided support for the validity of the second order model for middle school students (SRMR = .05, RMSEA = .04, CFI = .93, TLI = .93, and χ^2 (575, N = 301,513) = 319,553.46, $p < .01$).

Georgia School Climate Survey – Theoretical. A second, theoretical factor structure was developed with consideration of the aforementioned conceptual and statistical issues raised by school climate researchers (see, e.g., Konold et al., 2014; Marsh et al., 2009, 2012; Ramelow et al., 2015; Schweig, 2014; Wang & Degol, 2016). Items that were not specific to a student's experience in school were eliminated (e.g., the items included on the original GSCS's Character subfactor), and the remaining items were reorganized with respect to Wang and Degol's (2016) conceptualization and categorization of school climate. Seven subfactors of school climate were specified. This multifactor structure was analyzed using weighted least squares with mean and variance corrected (WLSMV) estimation in Mplus. Survey items were subsequently reduced based on CFA results and modifications indices with consideration of theoretical fit. Model fit statistics of the final model provided support for the validity of the 30-item, 7-factor model (see Table 2.2: RMSEA = .06, CFI = .96, TLI = .96, and $\chi 2$ (384, N = 259,778) = 354767.5, $p < 0.01$). Peer Relations and Adult Relations reflect the Quality of Relationships dimension of Wang and Degol's (2016) Community domain, while School Attachment and Respect for Diversity tap into the Connectedness and Respect for Diversity dimensions, respectively. Physical Environment incorporates aspects of both the Environmental and the Availability of Resources dimensions of the Institutional Environment domain, and Physical Safety reflects the Physical dimension of Wang and Degol's (2016) Safety domain. The Organizational Structure dimension of the theoretical GSCS does not clearly fit into just one of the domains specified by Wang and Degol (2016) but rather incorporates aspects of the Teaching & Learning dimension of the Academic domain and the Discipline & Order dimension of the Safety domain in that it captures both behavioral and academic expectations and practices.

Student Race/Ethnicity

The five racial/ethnic categories included in the GSHS were dummy coded, so that analyses included four race/ethnicity variables: "Asian/Pacific Islander" (1 0 0 0), "Black/African American" (0 1 0 0), "Hispanic/Latino" (0 0 1 0), and "Other" (0 0 0 1). "White/Caucasian" (0 0 0 0) was the referent group. The decision to use dummy coding procedures was due to the goal of identifying group differences. White/Caucasian was chosen as the referent category based on findings from prior school climate research that suggest students belonging to racial/ethnic minority groups report significantly lower perceptions of school climate than their White counterparts (see, e.g., Koth et al., 2008; Kuperminc et al., 1997; La Salle et al., 2016; White et al., 2014).

School Racial/Ethnic Composition

School racial/ethnic composition was measured using two variables: *Racial/Ethnic Diversity* and *Racial/Ethnic Share.* All school racial/ethnic composition data were gleaned from GOSA (2016a, 2016b).

School racial/ethnic diversity was measured using the Hirschman–Herfindahl Index (Hirschman, 1964) and calculated as follows:

$$1 - [(\%American\ Indian/Alaskan)^2 + (\%Asian/Pacific\ Islander)^2 + (\%Black)^2 + (\%Hispanic)^2 + (\%Multiracial)^2 + (\%White)^2] \quad (1)$$

The Hirschman–Herfindahl Index can be interpreted as the likelihood that two randomly drawn students do not belong to the same racial/ethnic group. Scores range can range from 0 to 1, which is reached when an infinite number of racial/ethnic categories are equally represented.

Thus, the value 0 means that there was no racial/ethnic diversity in the school, because all students identified with the same racial/ethnic group. Values that approach 1 represent a very high degree of diversity: e.g., students at that school identified with each racial/ethnic group at similar rates. The average *Racial/Ethnic Diversity* of schools was .47 (range = .02 to .77, *SD* = .18, *Mdn* = .52). This variable was skewed in a slightly negative direction (skewness = −.79).

Because the Hirschman–Herfindahl Index has been criticized for being ‘‘color-blind’’ (e.g. a school with 20% Black students and 80% White students obtains the same diversity score as a school with 20% White students and 80% Black students), the racial/ethnic share of the school was also included as a variable of school racial/ethnic composition (Dronkers & Van der Velden, 2013). This was calculated as: the proportion of students within each school who belong to racial/ethnic minority groups (i.e., the proportion of American Indian/Alaskan, Asian/Pacific Islander, Black, Hispanic, and Multiracial students). Scores can range from 0 to 1, with 0 indicating a school has no students from racial/ethnic minority groups and 1 indicating the entirety of a school's population belongs to racial/ethnic minority groups. While this calculation does not consider the effect of the proportion of specific racial/ethnic categories, it was chosen due to the disproportionate clustering of racial/ethnic groups within schools. In Georgia, Black and Hispanic/Latino students make up the majority of the student population after White students; and schools serving predominantly Black and/or Hispanic/Latino students suffer from a lack of resources (U.S. Department of Education Office for Civil Rights, 2012). The average *Racial/Ethnic Share* of schools was .58 (range = .02 to 1.00, *SD* = .28, *Mdn* = .56). This variable was normally distributed (skewness = .03) and was included in addition to *Racial/Ethnic Diversity* per recommendations by Dronkers and Van der Velden (2013). The two school-level racial/ethnic composition variables had a small correlation ($r = -.12$).

School Socioeconomic Status Composition

School SES composition refers to the proportion of students within each school that meet the state of Georgia's criteria for free-or-reduced lunch (FRL). That is, the proportion of students that falls into at least one of the following categories:

1. Lives in a family unit receiving Supplemental Nutrition Assistance Program (SNAP) food stamp benefits,
2. Lives in a family unit receiving Temporary Assistance for Needy Families (TANF) benefits, or
3. Identified as homeless, unaccompanied youth, foster, or migrant.

All SES composition data were gleaned from GOSA (2016a, 2016b). SES composition scores range from 0 to 1, where scores of 0 indicate that no students meet the above criteria (i.e. higher average SES) and scores of 1 indicate that all students meet the above criteria (i.e. lower average SES). The average *SES Composition* of schools was .35 (range = .01 to .85, SD = .28, *Mdn* = .35). This variable was approximately normally distributed (skewness = .34).

Data Analysis

The construct validity and invariance of the GSCS were examined in a stepwise approach (see below). *Research Question 1* was investigated within Steps 1 – 3, *Research Question 2* within Steps 4 and 5, and *Research Question 3* within Step 6. All analyses utilized the full data set. Factor analyses were conducted using robust maximum likelihood estimation (MLR) in Mplus Version 8 (Muthén, 1994). Multilevel CFA often uses MLR estimation because it does not assume normality and yields a robust chi-square (Kaplan et al., 2009; E. Kim et al., 2012). Due to the categorical nature of the indicators, analyses from Step 1 through Step 3a were also conducted using weighted least squares with mean and variance corrected (WLSMV) estimation, and results were compared to determine whether significant differences in interpretation arose across estimation methods. Because significant differences in model acceptability and selection did not arise (see Appendix A), MLR estimation was utilized to reduce computational demand.

Models were evaluated in terms of chi-square and alternative fit statistics, such as the comparative fit index (CFI; Bentler, 1990), Tucker-Lewis index (TLI; Tucker & Lewis, 1973), standardized root mean square residual (SRMR; Bentler, 1995), root mean square error of approximation (RMSEA; Steiger, 1990), Akaike information criterion (AIC; Akaike, 1974), Bayesian information criterion (BIC; Schwarz, 1978), and sample-size adjusted BIC (ssBIC; Sclove, 1987). When MLR is used for model estimation, the Yuan–Bentler chi-square test statistic is recommended for model comparison (for details, see, Yuan & Bentler, 2000). Due to chi-square's sensitivity to large sample sizes (Meade et al., 2008), decisions about changes in model fit were based predominantly on the alternative fit indices listed above. At each step, modification indices were examined and considered with school climate theory to determine if changes to the factor structure were warranted. When comparing models, a *decrease* in CFI of .01 and *increase* in RMSEA and SRMR of .015 and .03 indicate a *lack* of measurement invariance (Chen, 2007; Cheung & Rensvold, 2002). However, it is important to consider that these criteria were determined using single level models (E. Kim et al., 2016). Most fit indices assess overall fit and are not appropriate to evaluate fit at each level, particularly the between level because overall fit is predominantly influenced by the within level (E. Kim et al., 2016; Ryu & West, 2009). Thus, particular attention was paid to changes in SRMR at each level as it has been shown to detect misspecifications reasonably well (Hsu et al., 2015). In addition, changes in BIC and ssBIC were prioritized when comparing model fit, as these have been demonstrated to be optimal fit statistics when comparing multilevel models (E. Kim et al., 2015). Relatively smaller ssBIC and BIC values suggest better models.

Step 1: Conventional Factor Analysis

For a more in-depth explanation of the steps described below, see Appendix B. First, a conventional confirmatory factor analysis (CFA) was conducted to fit the proposed multifactor models to the total sample covariance matrix. The original GSCS model (see Table 2.1) and the proposed GSCS model (see Table 2.2) were specified. Results were analyzed to determine which model best fit the data and whether revisions to that model were indicated. Next, alternative factor structures of the best-fitting model were specified, including a unidimensional model, a second-order model, and a bifactor model (in which item variance is partitioned between a general school climate factor and specific factors, and the specific factors represent the shared variance of items not accounted for by the general school climate factor). Model fit statistics were compared to determine the best-fitting structure. Thus, the aim of *Step 1* is to identify what factor structure best explains the data (*Research Question 1a*) and to locate obvious model misspecifications using conventional methods (Muthén, 1994) before investigating the multilevel structure in *Research Question 1b.*

Step 2: Estimation of Variance and Reliability

The intra-class correlation (ICC) coefficients were calculated to examine the appropriateness of multilevel analysis (Hox 2010; Muthén, 1994, Muthén and Satorra, 1995). The ICC(1) indicates the proportion of the total variance in an item that can be explained by the school level. The ICC(2) coefficient is a measure of reliability that indicates the degree to which students within schools perceive school climate similarly. When ICC(1) coefficients are greater than .05 and ICC(2) coefficients are high (e.g., > .80), multilevel analysis is indicated.

Step 3: Multilevel Factor Analysis

Step 3a: Multilevel confirmatory factor analysis. An MCFA was conducted with the final factor identified in *Step 1* specified at both the within and between levels. Results were analyzed determine whether the GSCS factor structure fit the data at each level (*Research Question 1b*) or whether revisions to the factor structure were warranted.

Step 3b: Cross level measurement invariance. If the GSCS factor structure fits the data at each level, cross-level invariance procedures can be employed to ascertain whether school climate is being measured the same way at the individual-level and the school-level (*Research Question 1c*). Cross-level invariance procedures occur in a stepwise fashion. First, configural models are examined to determine if the factor structure and pattern of factor loadings is the same across levels (i.e. configural invariance). Because the factor structure of a single-level model is largely influenced by the within-school data, the between-schools level is investigated to explore whether alternative structures might better explain data. Next, the pattern and rank order correlations of factor loadings are examined across levels. If the factor structure and pattern of factor loadings is similar across levels, cross-level configural invariance can be assumed.

If support for configural invariance is found, the next step is to specify a metric model to test the hypothesis of equality of factor loadings across levels (i.e. *metric invariance*). If constraining the factor loadings to be the same across levels does not worsen the model fit, there is reasonable support that the constructs are similarly measured at the individual-level and the school-level. If it does worsen model fit, partial metric invariance can be explored by examining the modification indices to determine the source of non-invariance. When restrictions are placed on models (e.g., by constraining the factor loadings), Mplus identifies possible sources of misfit within the modification indices. Suggested factor loadings can be freed one at a time. Then, fit indices are examined to determine if model fit is improved (e.g., level-specific SRMR, ssBIC, BIC).

Step 3c: Cluster bias. Cluster bias was examined by constraining the final model from *Step 3b* so that the item residual variances at the between level were zero (Jak et al., 2013). If this constraint does not worsen model fit, it can be assumed that the survey measures school climate equally across schools (*Research Question 1d*). If the constraint does worsen model fit, modification indices can be examined to determine the presence of cluster bias in specific items (Jak et al., 2013). Suggested items can be freely estimated, one-by-one, until model fit is appropriate. If cluster bias is detected, school-level factors can be explored to determine if they account for the variability.

Step 4: Individual-level group measurement invariance

The final model from Step 3 was used to examine the invariance of the GSCS with respect to student-level variables (*Research Question 2a*). Testing the measurement invariance of multilevel models in relation to within-level groups is not feasible using MGCFA procedures because the group indicators (e.g., racial/ethnic groups) are crossed at the between level. To address this, the present study employed recommendations from Jak (2013)[1] and used ML MIMIC modeling procedures to test for invariance across student racial/ethnic groups. To investigate the assumption of invariance, a constrained model was constructed in which the student race/ethnicity grouping variable was added with direct effects specified on the corresponding within-level latent factors and constrained to zero on the within-level indicators.

[1]Procedures outline by Kim et al. (2015) to test for factor loading invariance using ML MIMIC modeling were initially employed but led to convergence issues. See *Appendix B* for more information.

The fit of this model was evaluated, and modification indices were examined to determine the presence of non-invariant items. If freeing a suggested direct effect does not improve model fit, intercept invariance can be assumed. If it does improve model fit, and if the direct effect is significant, it suggests the item's intercept is not invariant across groups.

Step 5: School-level measurement invariance

The final model from Step 4 was used to test for measurement invariance of the multilevel model in relation to school-level variables (*Research Question 2b*). Multilevel MIMIC modeling procedures were employed, which allow for invariance testing in reference to continuous variables (e.g. school racial/ethnic and SES composition). Between-level measurement invariance was estimated in a stepwise fashion, accounting for the effects of the within-level grouping factors (Jak, 2013). Multilevel MIMIC modeling invariance testing procedures were the same as described in *Step 4*, except paths were specified at the between level.

Step 6: Analyses of relationships

The final model from *Step 5* was examined to determine the relationships between the grouping variables and the overall school climate factor and its sub-factors (*Research Question 3*). Before structural relationships were interpreted, latent factor reliability estimates were calculated using procedures outlined in Rodriguez et al. (2016) to determine whether estimated relationships between latent factors and grouping variables could be considered trustworthy.

Results

Step 1: Conventional Factor Analysis

Conventional CFA results suggested that the theoretical GSCS model – in which items that did not directly reference the student's school were removed and items were reorganized in accordance with Wang and Degol's (2016) conceptualization of school climate – yielded slightly better fit than the original GSCS model (e.g., ΔCFI = .012, ΔTLI = .011, and ΔSRMR = −.009, see Table 2.3). Model fit statistics of the final multifactor model provided support for the validity of the 30-item, 7-factor model (CFI = .92, TLI = .91, RMSEA = .05, SRMR = .04, and $\chi 2$ (384, *N* = 259,778) = 210416.7, p < 0.01).

Next, alternative factor structures were specified. Results are shown in Table 2.3. A one-factor model, the most parsimonious of the models, yielded poor fit statistics (e.g., CFI = .63, RMSEA = .10). Next, a second order model was specified. Chi-square difference testing indicated that it did not significantly improve fit over the multifactor structure, and alternative fit-statistics suggested that it fit slightly worse than the multifactor model (ΔCFI = −.006, ΔTLI = −.003, ΔRMSEA = .001, and ΔSRMR = .004). Lastly, a bifactor model, in which item variance was partitioned between a general School Climate factor and the seven specific factors, was estimated. The fit of the bifactor model was a significant improvement over the second-order model, $\Delta\chi 2$(23, N = 259,778) = 29460.01, p < .001, and each of the alternative fit indices for this model suggested improved fit over the multifactor model (ΔCFI = .006, ΔTLI = .005, ΔRMSEA = –.001, and ΔSRMR = –.005). Thus, the bifactor model was identified as the best fitting GSCS factor structure (*Research Question 1a*) to be examined in *Research Question 1b* (see Figure 2.2).

The bifactor model provides the ability to investigate how item variance is partitioned between general and specific factors (Rodriquez et al., 2016). The GSCS items were generally good indicators of the general school climate factor, with standardized loadings ranging from .22 (item 32) to .67 (item 13), with only two loadings <.30. (items 32 and 37). When comparing those loadings with the item loadings for seven specific factors, slightly more total variance could be attributed to the general factor than to the specific factors (*M* = .50 & .46, respectively). However, this pattern varied by factor. For example, the adult relations and physical environment items all loaded more strongly on the general School Climate factor than on their respective specific factors, whereas the physical safety items all loaded more strongly on the specific Physical Safety factor than on the general School Climate factor. Item loadings ranged from .25 to .72 for School Attachment (*M* = .52), from .26 to .62 for Peer Relations (*M* = .41), from .45 to .60 for Adult Relations (*M* = .54), from .39 to .64 for Respect for Diversity (*M* = .52), from .31 to .50 for Physical Environment (*M* = .38), from .25 to .59 for Physical Safety (*M* = .45), and from .32 to .57 for Organizational Structure (*M* = .42). Omega coefficients for the bifactor model ranged from .94 for the School Climate general factor to .62 for the Physical Safety subfactor, with all other subfactor Omega coefficients greater than .73.

Step 2: Estimation of Variance and Reliability

The ICCs were calculated to examine the appropriateness of a multilevel analysis. The ICC(1) estimates ranged from .01 (Item 4) to .18 (Item 36), and 16 of the 30 items had ICC(1)s of .05 or greater. Thus, a substantial portion of the variability in student responses was attributable to school-level differences, supporting the use of MCFA. The ICC(2) estimates ranged from .90 (Items 4 and 9) to .99 (Items 28, 36, and 38), indicating a high level of agreement among students within the same school (i.e., students within schools tended to perceive school climate similarly) and suggesting the GSCS is a reliable estimate of school-constructs. Thus, support for the appropriate of the multilevel analyses proposed in *Research Questions 1-3* was found.

Step 3: Multilevel Factor Analysis.

Step 3a: Multilevel Confirmatory Factor Analysis

The final model from *Step 1* was specified at each level to test the measurement structure within and between groups (*Research Question 1b*). Note that a bifactor model with one general factor and seven specific factors and with item indicators for factors measured both within and between groups is a complex model. Nevertheless, the bifactor structure of the theoretical GSCS resulted in acceptable fit for the data at both the within and between levels (e.g., CFI = .92, TLI = .91, RMSEA = .03, SRMR Within = .04, and SRMR Between = .07; see Table 2.4). For the general School Climate factor, all item loadings, except for the school attachment items, were greater for the between-schools portion of the model (e.g., school level) than for the within-school portion (e.g., student level). Loadings ranged from −.04 (item 2) to .96 (item 89) for the between-schools model (*M* = .81) and ranged from .20 (items 32 and 37) to .65 (items 12, 13, and 14) for the within-school model (*M* = .48). The School Attachment factor loadings were lower for the within-school model, while, for all other specific factors, loadings were generally greater for the within-school model.

Step 3b: Cross Level Measurement Invariance

Because the GSCS factor structure fit the data at each level, cross-level invariance procedures were employed to ascertain whether school climate was being measured the same way at the individual-level and the school-level (*Research Question 1c*). First, configural invariance was examined. With the within-school model fully saturated, alternative models were specified at the between-schools level. Fit statistics provided support for the bifactor model over the unidimensional, multifactor, and second-order models (SRMR Between = .07, .10, .10, and .12, respectively).

With support for the factor structure across levels, the pattern of factor loadings was investigated. As previously noted, the loadings for the general School Climate factor were greater for the between-schools model for all items except the school attachment items (*M* = .48 and .81 for the within- and between-schools models, respectively). Spearman's rho rank order correlation for the School Climate factor was moderate (ρ = .67). School Attachment factor loadings were all greater for the between-schools model (ΔM =.32), but the rank order correlation was high (ρ = 1.0). The average factor loadings were similar across levels but somewhat greater for the within-school model (ΔM item loadings across levels = .02, .15, .22, .08, .07, and .15 for Peer Relations, Adult Relations, Respect for Diversity, Physical Environment, Physical Safety, and Organizational Structure, respectively). Except for Peer Relations (ρ = .10), the rank order correlations across levels were moderate to high for all factors (ρ = .40, .90, .60, 1.0, and .80 for Adult Relations, Respect for Diversity, Physical Environment, Physical Safety, and Organizational Structure, respectively). The most notable difference in the pattern of factor loadings across levels was that two items at the between level did not load onto their respective factors. Item 89 ("*The behaviors in my classroom allow teachers to teach so I can learn*.") did not significantly load on the Organizational Structure factor, nor did Item 2 ("*Most days I look forward to going to school*.") on the general School Climate factor. This suggests these items may not function the same way at the between level as they do the individual level.

In summary, the bifactor structure with one general factor and seven specific factors fit the data well at each level. The magnitude of School Climate and School Attachment factor loadings was somewhat greater for the between-schools model, but larger between-level factor loadings are expected due to the smaller sample size (Zyphur et al., 2008). The magnitude of loadings for the other specific factors was generally comparable across levels. Finally, the rank order correlations were moderate to high for all factors but Peer Relations, suggesting the patterns of factor loadings were similar. While the rank order correlation of Peer Relations was small, low rank correlations are not uncommon when *n* is small. Thus, overall support for configural invariance was concluded.

Because support for configural invariance was found, a metric model was specified to test the hypothesis of equality of factor loadings across levels (i.e. *metric invariance*). Constraining the factor loadings to be the same across levels worsened model fit, particularly for the between-schools model (ΔSRMR Within = .00, ΔSRMR Between = .20, ΔAIC = 2539.4, ΔBIC = 1995.0, and ΔSSBIC = 2160.3; see Table 2.4). Guided by modification indices, factor loadings were freed one at a time and results were analyzed to determine if model fit was improved. The final model specified 31 loadings free and 29 loadings constrained across levels. It demonstrated acceptable fit at each level (e.g., SRMR Within = .04 and SRMR Between = .07). The majority of freed loadings (20) were on the general School Climate factor, suggesting these items are not functioning the same way across levels. All of the loadings of the school attachment, physical environment, and physical safety items on the general School Climate factor were freed between levels, as were the majority of organizational structure and adult relations loadings. Notably, all of the School Climate loadings that were held constant between levels referenced student interactions (e.g., all of the peer relation items and items from other factors that mention student behaviors), suggesting that these items may have similar meanings in relation to School Climate across levels. School attachment items appeared to be more indicative of individual-level climate, while physical environment and safety items seemed to be more indicative of school-level climate.

In response to *Research Question 1c*, the bifactor structure of the theoretical GSCS demonstrated did not demonstrate cross-level invariance: general support for cross-level configural invariance was found, suggesting that the latent school climate factors are conceptually similar across levels; however, full invariance of factor loadings was not found, suggesting that the school climate factors cannot be interpreted the same way and are being measured differently across levels. Thus, the constructs may have disparate relationships to variables and outcome, and they should be modeled and explored separately.

Step 3c: Cluster Bias

Cluster bias was examined by constraining the final model from *Step 3b* so that the between-level residual variances were 0. This constraint worsened model fit (ΔSRMR Within = .00, ΔSRMR Between = .08, ΔAIC = 29231.7, ΔBIC = 28917.7, and ΔSSBIC = 29013.0; see Table 2.4). To determine the source of cluster bias, the modification indices were examined with consideration of the residual variance parameter estimates from the final metric model. The residual variances of suggested items were freed one-by-one and results were examined to determine if freeing the parameter led to improved model fit. The final model demonstrated acceptable fit across levels (e.g., SRMR Within = .04, and SRMR Between = .07) and indicated the presence of cluster bias in 24 items (i.e., freely estimated 24 between-level residual variances; see Table 2.5, Figure 2.3). Thus, in response to *Research Questions 1d*, support for invariance across schools was not found. The presence of cluster bias suggests that one or more school-level characteristics are violating measurement invariance at the between level. Thus, latent factors at the school level cannot be validly compared across diverse school populations. Students who possess similar perceptions of school climate but attend different schools may differ significantly with respect to their expected response on those items that contain cluster bias. Therefore, school-level factors (e.g., racial/ethnic composition, socioeconomic status composition, location) can be explored to determine if they account for the variability.

Step 4: Individual-Level Group Measurement Invariance

Before exploring whether school-level factors account for the cluster bias between schools, the final GSCS model from *Step 3* was explored to determine whether it functions similarly for students of various racial/ethnic groups (*Research Question 2a*). Multilevel MIMIC modeling procedures were employed, and an invariant model with direct effects from the dummy-coded race/ethnicity variables on each within-level latent factor freely estimated and with direct effects on the within-level indicators constrained to zero was specified. This model tested the assumption that within-school item intercepts are invariant across students of various racial/ethnic groups. Modification indices were examined to determine possible sources of noninvariance among item intercepts. Suggested direct effects from the racial/ethnic grouping variables to items were specified one at a time. If the direct effect was significant and model fit improved, evidence for non-invariance of that item's intercept was concluded and the direct effect was maintained in order to control for measurement error.

The final model demonstrated acceptable fit across levels (e.g., SRMR Within = .03, and SRMR Between = .08, see Table 2.6) and included 17 direct effects specified from the race/ethnicity variables on within-level indicators, suggesting that these 17 items may be functioning differently for students of different racial/ethnic groups. While several paths were specified, the standardized effects of student race/ethnicity on item intercepts were very small (i.e., magnitude of standardized path coefficient $\gamma < .1$ for many items). Racial/ethnic effects on item responses were largest (i.e., standardized $\gamma > .1$) for Item 4 ("*I feel successful at school.*"), Item 26 ("*I show courtesy to other students*."), and Item 37 ("*I have been involved in a fight at school.*"). Results suggested that, given similar perceptions of school climate, school attachment, peer relations, and physical safety, Latino/Hispanic students were slightly more likely to disagree with Item 4 and Item 26 (standardized $\gamma = -.10$ for each) than their White/European American counterparts, and African American students were somewhat more likely to agree with item 37 (standardized $\gamma = -.14$). Thus, in response to *Research Question 2a*, the GSCS did not demonstrate full invariance with respect to student racial/ethnic groups. Noninvariance of item intercepts suggests that something about belonging to certain racial/ethnic groups affects the way students respond to items in a systematic way, above and beyond the latent factors being measured. A partially invariant model was specified to account for this measurement error.

Step 5: School-Level Measurement Invariance

The final model from *Step 4* was used to explore whether school-level factors account for the cluster bias detected in *Step 3c*. Thus, measurement error with respect to student race/. ethnicity was controlled when exploring whether the GSCS functions similarly for schools with various racial/ethnic and socioeconomic compositions. To test for invariance with respect to school-level variables (*Research Question 2b*), ML MIMIC modeling procedures were employed. An invariant model was specified in which the school-level SES composition, racial/ethnic diversity, and racial/ethnic share variables were added to the model with direct effects specified on each latent factor in the between-schools model. As with *Step 4*, modification indices were examined to determine sources of noninvariance, and suggested direct effects from the racial/ethnic grouping variables to items were specified one at a time. If evidence for non-invariance was concluded, the direct effect was maintained in order to control for measurement error.

The final model demonstrated acceptable fit across levels (e.g., SRMR Within = .03, and SRMR Between = .04; see Table 2.6), and included 23 direct effects specified from school-level composition variables on 18 between-level indicators (see Table 2.6), suggesting that these 18 items may be functioning differently for schools with different SES and racial/ethnic compositions. Fourteen of the 18 items showed noninvariance with respect to only one school-level composition variable, three items (Items 6, 89, and 90) showed noninvariance with respect to two composition variables, one item (Item 37) showed noninvariance with respect to all three composition variables. The majority of item non-invariance was with respect to school SES composition and racial/ethnic share (e.g., ten items showed invariance with respect to each variable), and the SES composition of schools had the largest effect on item intercepts, suggesting climate comparisons across schools with different SES compositions may be particularly vulnerable to measurement error.

Items appeared to be more susceptible to bias at the school level than at the individual level. Item bias was largest for Item 1 ("*I like school*."), Item 6 ("*I get along with other students at school.*"), Item 9 ("*I have a group of friends at school that I have fun with and are nice to me*."), Item 19 ("*All students in my school are treated fairly, regardless of their appearance*."), Item 32 ("*I have felt unsafe at school or on my way to or from school*."), Item 37 ("*I have been involved in a fight at school*."), and Item 90 ("*Students are frequently recognized for good behavior*."). Results suggest that, given similar perceptions on corresponding latent factors, students in schools with greater proportions of low-SES students were more likely to disagree with Items 1 and 6 (standardized $\gamma = -.42$ and $-.36$, respectively), and they were more likely to agree with Items 19, 32, 37, and 90 (standardized $\gamma = .23, -.34, -.75$, and $.25$, respectively). Students in schools with a greater proportion of students from racial/ethnic minority groups were more likely to disagree with Items 9, 37, and 90 (standardized $\gamma = -.30, .38$, and $-.24$, respectively). Thus, in response to *Research Question 2b*, the GSCS did not demonstrate full invariance with respect to student racial/ethnic groups. Noninvariance of item intercepts suggests that something about school demographic compositions affects the way students respond to items in systematic way, above and beyond the latent factors being measured. The magnitude of school composition effects on item intercepts was greater at the school level than the individual level, suggesting that school composition variables may have a greater impact on noninvariance between schools than individual characteristics on noninvariance within schools. A partially invariant model was specified that accounts for measurement error at both the within- and between-levels (i.e., direct effects were specified from grouping variables onto noninvariant items).

Step 6: Analyses of Relationships

Before structural relationships were analyzed, the latent factor reliabilities were analyzed using the final measurement model from *Step 3*. Latent factor reliabilities estimate the quality of a latent factor's indicators and provide information on whether estimated relationships between a latent factor and other variables can be considered trustworthy. Results can be found in Table 2.7. Support for the reliability of the general School Climate factor was found across levels. At each level, the general School Climate factor demonstrated adequate internal reliability (e.g., $\omega > .90$). Further, it was found that the majority of variance in total scores at each level can be attributed to the general School Climate factor (e.g., ω_H and relative omega > .80). Thus, total scores at each level can be considered essentially unidimensional (Reise et al., 2013). Lastly, the general School Climate factor was well defined at both the within and between levels (e.g., $H > .90$), suggesting that estimated path coefficients between the general School Climate factors and other variables are trustworthy.

In contrast, school-level School Attachment was the only specific factor that demonstrated reliability sufficient for trustworthy interpretation. While the majority of other specific factors demonstrated adequate internal reliability (e.g., $\omega > .70$ for all specific factors except student-level Physical Safety), little common variance remained after partitioning out the variance for the general School Climate factor (e.g., $\omega_{HS} < .50$). In addition, these specific factors were not well-defined (e.g., $H < .70$), and the estimated path coefficients between them and other variables are not expected to replicate well across studies. In summary, latent factor reliability results provided support for the quality, replicability, and trustworthiness of the general School Climate factors across levels and the School Attachment factor at the between level. While the other specific factors were modeled to control for variance and to be true to the theoretical model of school climate and the best-fitting measurement model from *Step 3* (Coulacoglou & Saklofske, 2017), interpretation of structural relationships focused on the general School Climate factors and the school-level School Attachment factor.

The final GSCS structural model from *Step 5* – with each student- and school-level grouping variable included with appropriately specified paths to account for the noninvariance of item intercepts (as determined in *Steps 4* and *5*; see Figure 2.4) – was examined to determine the relationships between the grouping variables and the general School Climate factors and the school-level School Attachment factor (*Research Question 3*). Despite this model being complex, fit was acceptable: CFI = .92, TLI = .90, RMSEA = .03, SRMR Within = .03, SRMR Between = .04, and $\chi 2$ (840, N = 259,778) = 193043.6, $p < 0.01$. At the individual level, race/ethnicity had a small effect on School Climate. When accounting for the measurement noninvariance identified in *Step 4*, the largest effects of racial/ethnic group membership on perceptions of school climate were for Hispanic/Latino and African American students. Hispanic/Latino students reported somewhat more positive perceptions of individual-level School Climate than their European American counterparts (standardized $\gamma = .15$), while African

American students reported slightly more negative perceptions (standardized $\gamma = -.08$). The effects for Asian students and students of racial/ethnic groups not specified on the GSHS were significant but negligible (standardized $\gamma = .02$ and $-.04$, respectively).

At the school level, effects were larger. The proportion of racial/ethnic minorities within schools had the largest effect on the school-level School Climate factor. After accounting for the measurement noninvariance identified in *Step 5*, the racial/ethnic share of schools (e.g., proportion of the student body that identifies as non-white) had an inverse relationship to school climate. That is, as the share of students from racial/ethnic minority groups increased, perceptions of school-level School Climate decreased (standardized $\gamma = -.46$). The same pattern was found for the racial/ethnic diversity of schools (e.g., the number racial/ethnic groups present in the student body and how equally they are represented) and the socioeconomic composition. As racial/ethnic diversity and the share of students from low socioeconomic groups increased, perceptions of school-level School Climate decreased (standardized $\gamma = -.16$ and $-.28$, respectively). For school-level, residual School Attachment, this trend was reversed. After controlling for School Climate, as the share of students from racial ethnic minority groups and the share of students from low socioeconomic groups increased, perceptions of school-level School Attachment also increased (standardized $\gamma = .21$, and .75, respectively). Full results for the final model can be found in Table 2.8 and Table 2.9. However, the estimated direct effects of other specific factors should not be interpreted due to their low construct reliabilities (DeMars, 2013).

In summary, in response to *Research Question 3*, student race/ethnicity had a small relationship with individual-level School Climate, and school racial/ethnic and SES composition variables had a larger relationship with school-level School Climate and School Attachment. It is notable that, although the noninvariance of item intercepts appeared small at the student-level (e.g., magnitude standardized $\gamma < .15$), some of the relationships between race/ethnicity and individual-level School Climate were different (in magnitude and/or in direction) after accounting for the measurement noninvariance of items. For example, the standardized direct effect of Latino/Hispanic group membership on perceptions of School Climate changed from −.04 (*Step 4* invariant model) to .15 (final, partially invariant model). Similarly, the relationships between school composition variables and school-level latent factors were often different after accounting for the measurement noninvariance of items. For example, the effect of the SES composition of schools on school-level School Climate decreased in magnitude from −.49 (*Step 5* invariant model) to −.28 (final, partially invariant model). These findings suggest that failing to investigate and account for the noninvariance of items may lead to inaccurate conclusions about the relationships among school climate factors and variables of interest. With respect to the present data, if the noninvariance of item intercepts had not been accounted for, it may have inaccurately been concluded that Latino/Hispanic students possessed similar perceptions of School Climate as their White/European American counterparts when, in fact, results suggest that they possess somewhat more positive perceptions.

Discussion

The stakes attached to the accurate measurement of school climate are greater than ever. Subsequent to its inclusion in ESSA (2015), an increasing number of states are reporting school climate indicators within their accountability systems, and several statewide initiatives to measure it have been established (e.g., the *CAL-SCHLS System*: CDE, 2017; the *Delaware School Surveys*: Bear et al., 2016; and the *GSCS Suite*: GaDOE, 2016). Results from such surveys are often used for evaluation purposes and/or made available to the public (Bear et al., 2016; GaDOE, 2015). Thus, school climate measurement has real consequences, and it is critical that conclusions made using school climate data be valid. However, the complexity of school climate presents an array of challenges for accurate measurement, such as: the inclusion and/or exclusion of specific dimensions, the clustered nature of survey data, and the equality of surveys for diverse populations (Bear, 2016; Konold et al., 2014; Konold & Cornell, 2015; Wang & Degol, 2016; Zabek et al., 2017). Unfortunately, school climate initiatives often lack a conceptual model to justify decisions, fail to differentiate between the construct at the individual and school levels, and ignore its inherent hierarchical structure when evaluating it and determining the equality of surveys across diverse student and school populations (Konold et al., 2014; Marsh et al., 2009, 2012; Ramelow et al., 2015; Schweig, 2014; Wang & Degol, 2016). These trends are problematic and can result in incorrect conclusions and interpretations (E. Kim et al., 2012; Konold et al., 2014; Zyphur et al., 2008). The current study demonstrates the importance of using theory-grounded measurement development to investigate the multilevel nature of school climate data and the equality of school climate surveys for diverse student and school populations. It also provides evidence to support a general school climate factor for the GSCS at the individual and the school level and contributes findings regarding the survey's differential functioning across, and its relationship with, students of different races/ethnicities and schools of various racial/ethnic and SES compositions.

Measurement Development

Conventional CFA results (*Step 1*) favored the theoretical model of the GSCS that aligns with Wang and Degol's (2016) categorization of school climate and excludes items that are not specific to school experiences (e.g., "*Honesty is an important trait to me*"). In response to *Research Question 1a*, the data were best represented by a bifactor structure with one general school climate factor and seven specific factors (see Figure 2.2). Consistent with previous school climate research (e.g., Bear et al., 2011, 2015; La Salle et al., 2016; Zullig et al., 2015), the ICC estimates in the current study indicated that a meaningful proportion of variance in item responses was due to school-level characteristics (*Step 2*). This finding underscores the importance of utilizing multilevel modeling when measuring school climate. In practice, school climate is most often conceptualized and used as a school level construct; however, school climate surveys are typically validated using observed item responses at the individual level (e.g., Furlong et al., 2011; Johnson et al., 2007; La Salle et al., 2016; You et al., 2013). This results in measurement error, as it does not account for the variance unique to each level.

By employing a bioecological framework (Bronfenbrenner & Morris, 2006) and MCFA procedures, the current study was able to conceptualize, model, and interpret individual- and school-level climate simultaneously. The individual-level GSCS factors are conceptualized as *proximal processes* that reflect an individual's interactions with their school (see Figure 2.5). At the same time, the school-level GSCS factors represent *context* variables that estimate shared characteristics of a school's microsystem. Multilevel CFA procedures controlled for the measurement and sampling error associated with clustered data. Some researchers control for measurement error by centering their survey data around school means (e.g., Zullig et al., 2015). However, when the survey scores are then aggregated to be used as school-level constructs (e.g., Bear et al., 2011, 2014, 2015), error is still present, as this assumes cross-level invariance and does not account for the variability in number of individuals represented within different schools.

In response to *Research Question 1b*, multilevel CFA results found that the bifactor structure of the GSCS fit the data best at each level (*Step 3a*). Results indicated mixed support for the cross-level invariance of the GSCS (*Research Question 2b*). General support for configural invariance across levels was found (i.e., the factor structure and pattern of factor loadings were similar across levels; *Step 3b*), suggesting that the latent school climate factors are conceptually similar across levels. However, support for full cross-level invariance was not found (*Step 3c*). About half of the factor loadings demonstrated non-invariance across levels, suggesting that the school climate factors are cannot be interpreted the same way across levels and are being measured differently. While items reflecting student interactions seemed to be similarly indicative of climate across levels, school attachment items were more indicative of individual-level climate, and physical environment and safety items seemed to be more indicative of school-level climate. Thus, school-level climate cannot be interpreted as simply the aggregate of individual-level climate, and the constructs may have disparate relationships to variables and outcomes (Jak., 2019). This finding further emphasizes the importance of using multilevel methods to validate school climate surveys and to investigate the construct of school climate. When school climate surveys are validated using observed item responses (e.g., Furlong et al., 2011; Johnson et al., 2007; La Salle et al., 2016; You et al., 2013), cross-level invariance is assumed. Cross-level invariance is also assumed when researchers aggregate scores from surveys validated at the individual level (e.g., Bear et al., 2011, 2014, 2015). Incorrectly assuming cross-level invariance has notable consequences. Invalid dimensions and relationships may be concluded. The dimensions of school climate and its relationship with outcomes have a significant impact on policy and practice. By using the incorrect level of analysis to determine dimensions, qualities of school climate may be incorrectly targeted in policy and practice, or may be missed altogether (Schweig, 2014).

While the current study demonstrates the need for school climate surveys to be validated using multilevel methods, the use of such methods does not imply that the latent factors at each level can be compared across diverse individual and school populations while maintaining validity. Cluster bias results (*Step 3b*) indicated that the GSCS did not measure school climate equally across schools (*Research Question 1c*). Multilevel MIMIC modeling results (*Steps 4* and *5*) found that the individual-level GSCS did not demonstrate full invariance with respect to student racial/ethnic groups (*Research Question 2a*) and that the school-level GSCS did not demonstrate full invariance with respect to school racial/ethnic and SES compositions (*Research Question 2b*). At each level, over half of the items demonstrated some degree of noninvariance. Thus, these items and their associated latent factors have different meanings across certain racial/ethnic groups and school demographic compositions. Incorrectly assuming invariance may lead to invalid conclusions. Results from the current study revealed that school climate's relationship with some racial/ethnic groups and school composition demographics changed (in magnitude and/or in direction) after accounting for the noninvariance of items. These findings underscore the importance of ensuring the equality of the within- and between-level factor structures across diverse populations before making comparisons across individual subgroups or schools. Comparisons across schools with diverse SES compositions may be particularly vulnerable to measurement error. While invariance testing is becoming more frequent in school climate measurement development, it is still not universal (see, e.g., La Salle et al., 2016; White et al., 2014). This helps to ensure that mean differences and discovered relationships are not due to measurement error.

Results from the current study suggest that measurement bias in school climate surveys may be more pronounced at the school level than the individual level. The magnitude of school composition effects on item noninvariance was greater than that of student race/ethnicity. Presently, school climate invariance analyses occur almost exclusively at the individual level in relation to demographic characteristics such as gender, age, and race/ethnicity (see, e.g., Bear et al., 2011, 2014, 2015; Konold et al., 2014; Zullig et al., 2015). While investigating the invariance of school climate surveys in relation to individual-level variables is important, it is not sufficient in determining the quality of school climate surveys across schools. In addition, when invariance testing procedures do not account for the clustering of students within schools (see, e.g., Furlong et al., 2011; Johnson et al., 2007; You et al., 2013; Zabek et al., 2017), cross-level invariance is incorrectly assumed which may confound results. For example, if noninvariance in relation to student race/ethnicity is found, it may actually be an effect of school-level factors associated with schools of varying racial compositions. Thus, invariance with regard to individual- and school-level groups needs to be demonstrated using multilevel modeling techniques to ensure psychometrically valid comparisons of group differences are warranted.

Analyses of Relationships

General School Climate

By employing a bioecological framework (Bronfenbrenner & Morris, 2006) and MCFA procedures, the current study was able to model and interpret individual- and school-level climate simultaneously. In addition, by using ML MIMIC modeling, the current study was able to identify and control for item bias with respect to student race/ethnicity and school racial/ethnic and SES composition (*Steps 4 and 5*) before interpreting the relationships between these variables and school climate at each level (*Step 6*). In response to *Research Question 3*, multilevel MIMIC modeling results indicated that, when accounting for school context variables, individual-level school climate varied systematically as a function of *person* variables (i.e., student race/ethnicity); however, this relationship was very small (e.g., Latino/Hispanic students perceived school climate as somewhat more positive than their White/European American counterparts, while African American students perceived school climate as slightly less positive). At the school level, climate varied systematically as a function of other *context* variables (e.g., school composition). Students in schools that were more racially diverse and that had greater shares of students from low-income and racial/ethnic minority groups had more negative perceptions of school climate. The relationships between context variables and school-level climate was much stronger than the relationships between person variables and student-level climate, emphasizing the need for a conceptual and statistical framework that appropriately models these constructs, so phenomena at one level does not confound results at another.

Consistent with past research (see, e.g., Brault et al., 2014; Hopson, 2014; Newman et al., 1989), findings from the present study point to the potentially damaging effects of racial/ethnic segregation in schools. Schools serving greater proportions of racial/ethnic minority students had lower levels of overall climate. This effect was greater than the effect of school SES composition, suggesting that the income level of school populations cannot explain the more negative climates of schools serving larger numbers of students of color and students from ethnic minority groups. Thus, it is likely that there are other variables contributing to the relationship between climate and the racial/ethnic share of schools that should be explored in future studies.

The present study demonstrates the need for multilevel modeling procedures (e.g., MCFA) to account for the clustered nature of data when examining and interpreting school climate. Using conventional methods that did not account for the nested nature of previous studies have found that students identifying as racial/ethnic minorities perceive school climate less favorably than other students (see, e.g., Furlong et al., 2011; La Salle et al., 2016). Other studies have found that Black/African American students perceive school climate as less positive than students of other races and ethnicities (e.g., Furlong et al., 2011; Kuperminc et al., 1997;

White et al., 2014). In comparison, the present study found that identifying as Black/African American had a very small effect (i.e., magnitude of standardized $\gamma < .1$) on perceptions of school climate. However, the school-level race/ethnicity and SES composition variables had much larger effects. This finding suggests that racial/ethnic differences in perceptions of school climate may be due to confounding school-effects specific to schools that predominantly serve racial/ethnic minority populations. Thus, school climate researchers, educational practitioners, and policy makers must account for the multilevel nature of school climate data to avoid inaccurate interpretations of school climate data.

Findings from the current study also underscore the importance of level-specific invariance testing to ensure the equality of school climate surveys across diverse groups before examining and interpreting school climate surveys. Without first investigating measurement invariance, previous studies have found that there are no significant differences in school climate perceptions between Latino and White students (e.g., Kuperminc et al., 1997; White et al., 2014). In contrast, by using multilevel procedures (e.g., MCFA, ML MIMIC modeling) to account for the clustered nature of the data and the non-invariance of item intercepts, the present study found that Latino/Hispanic students perceived school climate as more positive than their White/European American counterparts. However, this relationship did not emerge until accounting for the measurement noninvariance of items. The standardized direct effect of Latino/Hispanic group membership on perceptions of individual-level School Climate changed from −.04 (when noninvariance was not taken into account) to .15 (when controlling for noninvariance). Not accounting for noninvariance (e.g., specifying direct effects from the grouping variables onto the biased items) may have led to the interpretation – similar to that of the studies described above – that Latino/Hispanic and White/European students perceive school climate similarly. Thus, there are practical implications of not accounting for item bias – even if that bias appears to be negligible – as relationships between student racial/ethnic group membership and perceptions of school climate may be due to measurement noninvariance at the individual level.

School-Level Attachment

Findings from the present study suggest that, at the school level, school attachment and school climate had disparate relationships with *context* variables (e.g., school composition). It is important to interpret this finding with an understanding that specific factors within a bifactor model are not equivalent to factors within multifactor or second order models. In the bifactor model, specific factors are residuals relative to the general factor. They are *independent from* the general factor and represent “variance common to a group of items beyond the factor measured by the scale as a whole” (DeMars, 2013, p. 355). Here, the specific factor represents school-level attachment *above and beyond* the attachment reflected in the school-level climate factor. Results from this factor reveal patterns in school-level attachment after controlling for the general school climate factor (and the perceptions of school attachment that contribute to it). Findings from the present study demonstrate that, after controlling for general school climate at the school-level, schools with greater proportions of students from low-income and racial/ethnic minority groups had more positive perceptions of school attachment. Thus, while schools with greater proportions of students from low-income and racial/ethnic minority groups had more negative school climates, they had relatively higher levels of residual school attachment (i.e., school attachment over and above that accounted for in the school climate factor). Attending schools where students who are generally well attached to school has been shown to be a protective factor regardless of an individual student's feelings of school attachment (Henry & Slater, 2007). Thus, future research should explore this finding further.

The Georgia School Climate Survey

The present study found that the bifactor structure of the theoretical GSCS fit the data best at both the individual and school levels. However, cross-level invariance was not found. Thus, the individual- and school-level constructs should be modeled and explored separately when utilizing the GSCS to explore climate. Latent factor reliability results provided support for the quality, replicability, and trustworthiness of the general School Climate factors across levels. However, except for the School Attachment at the between level, the GSCS specific factors were not well defined and did not demonstrate reliability sufficient for trustworthy interpretation. While the other specific factors were modeled within the current study in order to control for variance and to be true to the theoretical model of school climate and the best-fitting measurement model from study results, revisions to the GSCS may be warranted. To be true to the theoretical model of school climate as a multidimensional construct, future research may focus on strengthening the reliability of specific factors. Conversely, future research may use factor reliability results from the current study to create a unidimensional construct at each level. Results indicated that the majority of variance in total scores at each level could be attributed to the general School Climate factor (e.g., ω_H and relative omega > .80). Thus, total scores at each level could be considered essentially unidimensional (Reise et al., 2013).

Measurement invariance analyses found evidence for the partial invariance of the GSCS at each level with respect to student race/ethnicity and school racial/ethnic and SES composition. Four items were invariant across levels: Item 12 ("*Adults in this school treat all students with respect.*"), Item 13 ("*All students are treated fairly by the adults in my school.*"), Item 29 ("*My* *textbooks are up to date and in good condition.*"), and Item 92 ("*I know what to do if there is an emergency at my school.*"). Thus, these items may be prioritized for inclusion in future iterations of the GSCS. Of the 26 items that showed evidence of non-invariance at either level, future revisions may choose to also include items that showed the most negligible levels of noninvariance (e.g., standardized direct effects of < .06): Item 11 ("*Teachers treat me with respect*"), Item 14 ("*Teachers treat all students fairly.*"), Item 17 ("*Students show respect to other students regardless of their academic ability.*"), Item 18 ("*Students at this school are treated fairly by other students regardless of race, ethnicity, or culture.*"), Item 28 ("*My school building is well maintained.*"), Item 38 ("*I have observed a fight at school.*), and Item 88 ("*The behaviors in my classroom allow teachers to teach so I can learn.*"). Notably, over half of the items that showed no-to-little bias (6 of 11) referenced treating others fairly or with respect. This suggests that treating others fairly or with respect has similar meaning with respect to school climate across racial/ethnic groups and school racial/ethnic and SES compositions. Future iterations of the GSCS should consider revising or removing the items that showed the highest levels of noninvariance. Item 37 ("*I have been involved in a fight at school.*") demonstrated the most bias. It demonstrated noninvariance with respect to all four demographic variables and showed the largest standardized γ at each level. This suggests that being involved in a fight at school has very different meanings in relation to school climate across racial/ethnic groups and school racial/ethnic and SES compositions. Thus, it should be removed from future iterations of the GSCS.

The current study utilized an existing database (i.e., the GSHS data) to investigate the GSCS. This database included both self- and school-referent items. While many school climate surveys include both self- and school-referent items (see, e.g., Bear et al., 2011, Furlong et al., 2011; Johnson et al., 2007; La Salle et al., 2016; White et al., 2014; Zullig et al., 2015), researchers have argued that the meaning of a construct at the individual- and school-level may differ depending on whether it is based on self- or group-referent items, as self-referent items are designed to reflect individual-level constructs and school-referent items are designed to reflect school-level constructs (see, e.g., Bliese, 2000, Chan, 1998, Lüdtke et al., 2008; Marsh et al., 2009; Morin et al., 2014). Still, both are typically influenced by factors at both levels. The ICC(1) results within the current study found that 5% or more of the variability in responses was attributable to school-level differences for 16 items. Of these 16 items, all but five were school referent. Of the 14 items for which less than 5% of the variability in responses was attributable to school-level differences, all but three were self-referent. Consistent with other research (e.g., see Marsh et al., 2009; Morin et al., 2014), this finding suggests that self-referent school climate items reflect predominantly individual-level constructs and that school-referent items better reflect school-level constructs. However, responses to both types of items are most strongly influenced by individual differences. Still, future researchers may consider creating different sets of items for school climate surveys and including only self-referent items in surveys intended to measure the individual-level and only school-referent items in surveys intended to measure the school-level.

Limitations and Future Directions

Several limitations must be kept in mind when interpreting these results. First, the current study relied student self-report data, which are vulnerable to respondent biases. However, individuals' perceptions are critical for understanding behavior and utilizing them allowed for the present study to examine similarities and differences in perceptions across different groups of students. Future research may consider comparing results of school climate surveys to more objective measures of school climate (e.g., observation checklists).

Second, all participants were public middle school students in Georgia. Nevertheless, the sample was very diverse in that it included students and schools throughout urban, suburban, and rural parts of the state. Students were racially/ethnically diverse, and schools had varying degrees of racial/ethnic and SES compositions. Still, future research should explore the multilevel nature of school climate and the equality of its measurement using the perceptions of other stakeholders (e.g., elementary and high school students, parents, teachers) from other locations.

Third, the survey included a combination of self- and school-referent items, which researchers have argued may change the meaning of constructs at each level. However, this allowed for a meaningful examination of how each type of item functions across levels (e.g., the variability in item responses due to individual- and school- characteristics). Future research may further explore the functioning of items across levels or may consider creating level-specific item sets when exploring school climate.

Fourth, data were predominantly treated as continuous using MLR estimation to reduce computational demand. MCFA often uses MLR estimation because it does not assume normality and yields a robust chi-square (Kaplan et al., 2009; E. Kim et al., 2012). Still, to support the validity of this approach, *Steps 1-3a* were also conducted using WLSMV estimation (see Appendix A). Results were similar across methods; thus, MLR was used for subsequent analyses. Future researchers may consider comparing in more depth how MLR and WLSMV estimation perform when conducting MCFA.

Fifth, the measurement invariance of the GSCS was only explored with respect to student race/ethnicity and school level race/ethnicity and SES composition. Due to the complexity of the model explored, the present study limited the number of demographic variables explored. The individual and school characteristics investigated were chosen due to the high degree of racial/ethnic segregation across schools. Thus, these characteristics allowed for the present study to examine how the disproportionate clustering within schools may impact the results of validation analyses. Still, it is important that future research further explore the invariance of school climate surveys in relation to other individual and school characteristics (e.g., student family income, disability status, and language status; school type, geographic location, and funding).

Sixth, the procedures outlined by E. Kim et al. (2015) to test for factor loading invariance using ML MIMIC modeling were initially employed but led to convergence issues (see Appendix B for more information). Other studies have using this procedure have also resulted in convergence problems and in biased parameter estimates when the indicator is not normal (Bagheri et al., 2018; Cham et al., 2012); and researchers have argued that intercept noninvariance is more important than factor loading noninvariance, because noninvariance of factor loadings often evens out at the scale level (Huang et al., 2011). Thus, only item intercept invariance was tested in the present study, using ML MIMIC modeling procedures to build a model that controls for intercept noninvariance.

Finally, the functioning of the GSCS was only investigated in relation to demographic *person* and *context* variables, not *time* variables (Bronfenbrenner & Morris, 2006). This allowed for focused attention on the functioning of the survey across student and school characteristics. Future research should examine how individual- and school-level climate interact with person and context variables to impact outcomes of interest (e.g., achievement, mental health) over time.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. http://dx.doi.org/10.1109/tac.1974.1100705

Bagheri, Z., Akbari, M., Jafari, P., & Huedo-medina, T. B. (2018). Assessing the measurement invariance of the RCADS-25 questionnaire across gender and child–parent dyads in the presence of multilevel data. *Quality of Life Research, 28*(3), 829–840. http://dx.doi.org/10.1007/s11136-018-2049-3

Battistich, V., Solomon, D., Kim, D., Watson, M., & Schaps, E. (1995). Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance: A multilevel analysis. *American Educational Research Journal, 32*, 627–658. http://dx.doi.org/10.2307/1163326

Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2011). Delaware School Climate Survey-Student: Its factor structure, concurrent validity, and reliability. *Journal of School Psychology, 49*, 157–174. http://dx.doi.org/10.1016/j.jsp.2011.01.001

Bear, G., Yang, C., Harris, A, Mantz, L, Hearn, S., Boyer, D. (2016, November). *Technical manual for the Delaware School Survey: Scales of school climate; bullying victimization; student engagement; positive, punitive, and social emotional learning techniques; and social and emotional competencies*. Retrieved from http://wh1.oet.udel.edu/pbs/school-climate/administration-of-survey/

Bear, G. G., Yang, C., & Pasipanodya, E. (2015). Assessing school climate: Validation of a brief measure of the perceptions of parents. *Journal of Psychoeducational Assessment, 33*(2), 115–129. http://dx.doi.org/10.1177/0734282914545748

Bear, G. G., Yang, C., Pell, M., & Gaskins, C. (2014). Validation of a brief measure of teachers' perceptions of school climate: relations to student achievement and suspensions. *Learning Environments Research, 17*(3), 339–354. http://dx.doi.org/10.1007/s10984-014-9162-1

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107*, 238–246. http://dx.doi.org/10.1037/0033–2909.107.2.238

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bliese, P. D. (2000). Within-group agreement, nonindependence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bottoni, G. (2016). A multilevel measurement model of social cohesion. *Social Indicators Research, 136*(3), 835–857. http://dx.doi.org/10.1007/s11205-016-1470-7

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

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist 32*(7), 513–531.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 1: Theoretical models of human development* (6th ed., pp. 793 – 828). New York: Wiley.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. Journal of Personality, 80(4), 796-846. http://dx.doi.org/10.1111/j.1467-6494.2011.00749.x

California Department of Education. (2017). *Assessing social-emotional learning and health, 2016-17*. Retrieved from http://cal-schls.wested.org

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with non-normal observed data: A comparison of four approaches. *Multivariate Behavioral Research*, *47*(6), 840–876. http://dx.doi.org/10.1080/00273171.2012.732901

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*, 234–246. http://dx.doi.org/10.1037//0021-9010.83.2.234

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. http://dx.doi.org/10.1080/10705510701301834

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of personality and social psychology*, *95*(5), 1005–1018. http://dx.doi.org/10.1037/a0013193

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. http://dx.doi.org/10.1207/s15328007sem0902_5

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

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

Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental health agencies. *Journal of Emotional and Behavioral Disorders, 19*, 27–40. http://dx.doi.org/10.1177/1063426610365879

DeMars, C. E. (2013). S tutorial on interpreting bifactor model scores. *International Journal of Testing, 13*(4), 354–378. http://dx.doi.org/10.1080/15305058.2013.799067

Dronkers, J., & Van der Velden, R. (2013). Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using PISA data. In M. Windzio (Ed.), *Integration and inequality in educational institutions* (pp. 71–98). New York, NY: Springer. http://dx.doi.org/10.1007/978-94-007-6119-3_4

Every Student Succeeds Act (ESSA) of 2015, Public Law No: 114-95. December 10, 2015.

Frankenberg, E., Lee C., & Orfield, G. (2003). *A Multiracial Society with Segregated Schools*. Cambridge, MA: Civil Rights Project, Harvard University.

Freiberg, H. J., & Stein, T. A. (1999). Measuring, improving and sustaining healthy learning environments. In H. J. Freiberg (Ed.), *School climate: Measuring, improving and sustaining healthy learning environments* (pp. 11–29). Philadelphia, PA: Falmer Press.

Furlong, M. J., Greif, J. L., Bates, M. P., Whipple, A. D., & Jimenez, T. C. (2005). Development of the California School Climate and Safety Survey—Short form. *Psychology in the Schools, 42*, 137–149. http://dx.doi.org/10.1002/pits.20053

Furlong, M. J., O'Brennan, L. M., & You, S. (2011). Psychometric properties of the Add Health School Connectedness Scale for 18 sociocultural groups. *Psychology in the Schools, 48*(10), 986–997. http://dx.doi.org/10.1002/pits.20609

Georgia Department of Education. (2016). *Georgia Student Health Survey 2.0: School Climate Survey Suite*. Available by request from GaDOE.

Georgia Department of Education. (2015, May 4). *2014 School Climate Star Ratings data calculation guide for principals and district users*. Retrieved from http://www.gadoe.org/External-Affairs-and-Policy/Policy/Pages/School-Climate.aspx

Georgia's Governor's Office of Student Achievement. (2016a). *Enrollment by subgroup programs: 2015-2016* [Data file]. Retrieved from https://gosa.georgia.gov/downloadable-data-files-0

Georgia's Governor's Office of Student Achievement. (2016b). *Directly certified (school level): 2015-2016* [Data file]. Retrieved from https://gosa.georgia.gov/downloadable-data-files-0

Griffith, J. (2000). School climate as group evaluation and group consensus: Student and parent perceptions of the elementary school environment. *Elementary School Journal, 101*, 35–61. http://dx.doi.org/10.1086/499658

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*, 373–388. http://dx.doi.org/10.1007/BF02294440

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*, 60–87. http://dx.doi.org/10.3102/0162373707299706

Henry, K. L., & Slater, M. D. (2007). The contextual effect of school attachment on young adolescents' alcohol use. *Journal of School Health, 77*(2), 67-74. http://doi.org/10.1111/j.1746-1561.2007.00169.x

Hirschman, A.O., 1964. The paternity of an index. *The American Economic Review 54*(5), 761–762.

Hopson, L. M., Lee, E., & Tang, N. (2014). A multi-level analysis of school racial composition and ecological correlates of academic success. *Children and Youth Services Review, 44*, 126–134. http://doi.org/10.1016/j.childyouth.2014.05.026

Hox, J.J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Hsu, H.-Y., Kwok, O., Lin, J. H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A monte carlo study. *Multivariate Behavioral Research, 50*(2), 197–215. http://dx.doi.org/10.1080/00273171.2014.977429

Huang, I.-C., Leite, W. L., Shearer, P., Seid, M., Revicki, D. A., & Shenkman, E. A. (2011). Differential item functioning in quality of life measure between children with and without special health-care needs. *Value in Health, 14*(6), 872–883. http://dx.doi.org/10.1016/j.jval.2011.03.004

Jak, S. (2013). Cluster bias: Testing measurement invariance in multilevel data.

Jak, S. (2018). Cross-level invariance in multilevel factor models. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(4), 607–622. http://dx.doi.org/10.1080/10705511.2018.1534205

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling:*

A Multidisciplinary Journal, *20*(2), 265–282. http://dx.doi.org/10.1080/10705511.2013.769392

Johnson, B., Stevens, J. J., & Zvoch, K. (2007). Teachers' Perceptions of School Climate: A Validity Study of Scores From the Revised School Level Environment Questionnaire. *Educational and Psychological Measurement*, *67*(5), 833–844. https://doi.org/10.1177/0013164406299102

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, *8*, 325–352.

Kaplan, D., Kim, J. S., & Kim, S. Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 592–612). Thousand Oaks, CA: Sage.

Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research*, *51*(6). https://doi.org/10.1080/00273171.2016.1228042

Kim, E. S., Kwok, O., & Yoon, M. (2012). Testing factorial invariance in multilevel data: A Monte Carlo study. *Structural Equation Modeling*, *19*, 250–267. http://dx.doi.org/10.1080/10705511.2012.659623

Kim, E. S., Yoon, M., Wen, Y., Luo, W., & Kwok, O. M. (2015). Within-level group factorial invariance with multilevel data: Multilevel factor mixture and multilevel MIMIC models. Structural Equation Modeling: A Multidisciplinary Journal, 1–14. http://dx.doi.org/10.1080/10705511.2014.938217

Kim, H. Y., Schwartz, K., Cappella, E., & Seidman, E. (2014). Navigating middle grades: Role of social contexts in middle grade school climate. *American Journal of Community Psychology*, *54*(1–2), 28–45. http://dx.doi.org/10.1007/s10464-014-9659-x

Konold, T., & Cornell, D. (2015). Multilevel multitrait–multimethod latent analysis of structurally different and interchangeable raters of school climate. *Psychological Assessment*, *27*(3), 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

Konold, T., Cornell, D., Shukla, K., & Huang, F. (2017). Racial/ethnic differences in perceptions of school climate and its association with student engagement and peer aggression. *Journal of Youth & Adolescence, 46*(6), 1289–1303. http://dx.doi.org/10.1007/s10964-016-0576-1

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. Journal of *Educational Psychology, 100*(1), 96–104. http://dx.doi.org/10.1037/0022-0663.100.1.96

Kuperminc, G. P., Leadbeater, B. J., Emmons, C., & Blatt, S. J. (1997). Perceived school climate and difficulties in the social adjustment of middle school students. *Applied Developmental Science, 1*, 76–88. http://dx.doi.org/10.1207/s1532480xads0102_2

La Salle, T. P., Zabek, F., & Meyers, J. (2016). Elementary student perceptions of school climate and associations with individual and school factors. *School Psychology Forum, 10*(1), 55-65.

Lee, E., Reynolds, K. J., Subasic, E., Bromhead, D., Lin, H., Marinov, V., & Smithson, M. (2017). Development of a dual school climate and school identification measure–student (SCASIM-St). Contemporary Educational Psychology, 49, 91–106. http://dx.doi.org/10.1016/j.cedpsych.2017.01.003

Longford, N. T., & Muthén, B. O. (1992). Factor analysis for clustered observations. *Psychometrika*, *57*(4), 581–597. http://dx.doi.org/10.1007/bf02294421

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. http://dx.doi.org/10.1037/a0012869

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects. *Educational Psychologist, 47*(2), 106–124. http://dx.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. http://dx.doi.org/10.1080/00273170903333665

McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum Associates.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*(3), 259–284. http://dx.doi.org/10.1037/1082-989X.10.3.259

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. http://dx.doi.org/10.1007/bf02294825

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

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585. http://dx.doi.org/10.1007/bf02296397

Muthén, B.O. (1994). Multilevel covariance structure-analysis. *Sociological Methods & Research, 22*, 376–398. http://dx.doi.org/10.1177/0049124194022003006

Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 15–40). New York, NY: Taylor & Francis.

Muthén, B.O. & Satorra, A. (1995). Complex ample data in structural equation modeling. *Sociological Methodology, 25*, 267–316. http://dx.doi.org/10.2307/271070

National School Climate Council. (2007). *The School Climate Challenge: Narrowing the gap between school climate research and school climate policy, practice guidelines and teacher education policy*. Retrieved from http://www.schoolclimate.org/climate/documents/school-climate-challenge.pdf

Newmann, F. M., Rutter, R. A., & Smith, M. S. (1989). Organizational factors that affect school sense of efficacy, community, and expectations. Sociology of Education, 62(4), 221. http://dx.doi.org/10.2307/2112828

Orfield, G., & Frankenberg, E. (2014). Increasingly Segregated and Unequal Schools as Courts Reverse Policy. Educational Administration Quarterly, 50(5), 718–734. http://dx.doi.org/10.1177/0013161x14548942

Ramelow, D., Currie, D., & Felder-Puig, R. (2015). The assessment of school climate: Review and appraisal of published student-report measures. *Journal of Psychoeducational Assessment, 3*3(8), 731–743. http://dx.doi.org/10.1177/0734282915584852

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129–140. http://doi.org/10.1080/00223891.2012.725437

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357. http://dx.doi.org/10.2307/2087176

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. http://dx.doi.org/10.1037/met0000045

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 583–601. http://dx.doi.org/10.1080/10705510903203466

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464 http://dx.doi.org/10.1214/aos/1176344136

Schweig, J. (2014). Cross-level measurement invariance in school and classroom environment surveys: Implications for policy and practice. *Educational Evaluation and Policy Analysis*, *36*(3), 259–280. http://dx.doi.org/10.3102/0162373713509880

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333–343. http://dx.doi.org/10.1007/bf02294360

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches. In F. J. Van de Vijver, D. A. Van Hemert, & Y. H. Poortinga (Eds.), Multilevel analysis of individuals and cultures (pp. 93–119). Mahwah: Lawrence Erlbaum Associates.

Sirotnik, K. A. (1980). Psychometric implications of the unit-of-analysis problem (with examples from the measurement of organizational climate). *Journal of Educational Measurement, 4*, 245–282. http://dx.doi.org/10.1111/j.1745-3984.1980.tb00831.x

Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239. http://dx.doi.org/10.1111/j.2044-8317.1974.tb00543.x

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173–180. http://dx.doi.org/10.1207/s15327906mbr2502_4

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

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10. http://dx.doi.org/10.1007/bf02291170

Tudge, J. R. H., Mokrova, I., Hatfield, B. E., & Karnik, R. B. (2009). Uses and misuses of Bronfenbrenner's bioecological theory of human development. *Journal of Family Theory & Review, 1*, 198–210. http://dx.doi.org/10.1111/j.1756-2589.2009.00026.x

U.S. Census Bureau. (2011, March). *Overview of race and Hispanic origin: 2010*. Retrieved January 8, 2015 from http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf

U.S. Census Bureau. (2012, May). Most children younger than age 1 are Minorities, Census Bureau Reports. Retrieved January 8, 2015 from https://www.census.gov/newsroom/releases/archives/population/cb12-90.html

Vieno, A., Perkins, D. D., Smith, T. M., & Santinello, M. (2005). Democratic school climate and sense of community in school: A multilevel analysis. *American Journal of Community Psychology, 36*, 327–341. http://dx.doi.org/10.1007/s10464-005-8629-8

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

White, N., La Salle, T., Ashby, J. S., & Meyers, J. (2014). A brief measure of adolescent perceptions of school climate. *School Psychology Quarterly, 29(3),* 349–359. http://dx.doi.org/10.1037/spq0000075

You, S., O'Malley, M. D., & Furlong, M. J. (2013). Preliminary development of the Brief–California School Climate Survey: dimensionality and measurement invariance across teachers and administrators. *School Effectiveness and School Improvement*, *25*(1), 153–173. http://dx.doi.org/10.1080/09243453.2013.784199

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*(1), 165–200. http://dx.doi.org/10.1111/0081-1750.00078

Zabek, F., Rice, K. G., Davis, D. E., La Salle, T. P., and Meyers, J. (2017). *A grateful climate: Measurement invariance, demographic differences, and the relationship between school* *climate and gratitude* [Unpublished manuscript]. Department of Counseling and Psychological Services, Georgia State University.

Zullig, K. J., Collins, R., Ghani, N., Hunter, A. A., Patton, J. M., Huebner, E. S., & Zhang, J. (2015). Preliminary development of a revised version of the School Climate Measure. *Psychological Assessment, 27*(3), 1072–1081. http://dx.doi.org/10.1037/pas0000070

Zyphur, M., Kaplan, S., & Christian, M. (2008). Assumptions of cross-level measurement and structural invariance in the analysis of multilevel data: problems and solutions. *Group Dynamics: Theory, Research, and Practice, 12*, 127–140.

Table 2.1
Original Georgia School Climate Scale Items and Subfactors

School Connectedness
1. I like school.
2. Most days I look forward to going to school.
3. I feel like I fit in at my school.
4. I feel successful at school.
5. I feel connected to others at school.

Character
6. I treat other students fairly.
7. Doing the right thing is important to me.
8. I am open towards different opinions and perspectives.
9. I believe in helping others.
10. Honesty is an important trait to me.
11. I show courtesy to other students.

Physical Environment
12. My school building is well maintained.
13. My textbooks are up to date and in good condition.
14. Teachers in my school keep their classrooms clean and organized.
15. Students in my school take pride in keeping our school building (e.g. bathrooms, classrooms, lockers) in good condition.

Adult Social Support
16. Teachers treat me with respect.
17. Adults in this school treat all students with respect.
18. All students are treated fairly by the adults in my school.
19. Teachers treat all students fairly.

Peer Social Support
20. I get along with other students at school.
21. I know a student at my school that I can talk to if I need help (e.g., homework, class assignments, projects).
22. Students in my school are welcoming to new students.

Cultural Acceptance
23. Students at my school treat each other with respect.
24. Students treat one another fairly.
25. Students show respect to other students regardless of their academic ability.
26. Students at this school are treated fairly by other students regardless of race, ethnicity, or culture.
27. All students in my school are treated fairly, regardless of their appearance.

Order and Discipline
28. I feel my school has high standards for achievement
29. My school has clear rules for behavior
30. The behaviors in my classroom allow teachers to teach so I can learn
31. Students are frequently recognized for good behavior
32. I know an adult at school that I can talk with if I need help

Safety
33. I have felt unsafe at school or on my way to or from school
34. I have worried about students hurting me
35. I have been concerned about my physical safety at school
36. Students at my school fight a lot.

Table 2.2

Theoretical Georgia School Climate Scale Items and Subfactors

School Attachment
1. I like school.
2. Most days I look forward to going to school.
4. I feel successful at school.
Peer Relationships
6. I get along with other students at school.
7. I know a student at my school that I can talk to if I need help (e.g., homework, class assignments, projects).
9. I have a group of friends at school that I have fun with and are nice to me.
20. I treat other students fairly.
26. I show courtesy to other students.
Adult Relationships
11. Teachers treat me with respect.
12. Adults in this school treat all students with respect.
13. All students are treated fairly by the adults in my school.
14. Teachers treat all students fairly.
Respect for Diversity
15. Students at my school treat each other with respect.
16. Students treat one another fairly.
17. Students show respect to other students regardless of their academic ability.
18. Students at this school are treated fairly by other students regardless of race, ethnicity, or culture.
19. All students in my school are treated fairly, regardless of their appearance.
Physical Environment
28. My school building is well maintained.
29. My textbooks are up to date and in good condition.
30. Teachers in my school keep their classrooms clean and organized.
31. Students in my school take pride in keeping our school building (e.g. bathrooms, classrooms, lockers) in good condition.
Safety
32. *I have felt unsafe at school or on my way to or from school. (Reversed)*
36. *Students at my school fight a lot. (Reversed)*
37. *I have been involved in a fight at school. (Reversed)*
38. *I have observed a fight at school. (Reversed)*
Organizational Structure
87. I feel my school has high standards for achievement
88. My school has clear rules for behavior
89. The behaviors in my classroom allow teachers to teach so I can learn
90. Students are frequently recognized for good behavior
92. I know what to do if there is an emergency at my school.

Table 2.3
Fit Statistics for Step 1: Conventional CFA

	χ^2	df	CFI	TLI	RMSEA	SRMR	AIC	BIC	SSBIC
GSCS Models									
Original GSCS	301037.3*	566	.909	.899	.045	.051	19741069.6	19742493.2	19742061.0
Theoretical GSCS	210416.7*	384	.921	.910	.046	.042	16923663.2	16924825.1	16924472.3
Theoretical GSCS Alternative Models									
One Factor	978444.6*	405	.631	.604	.096	.077	17906405.8	17907347.9	17907061.9
Multifactor	210416.7*	384	.921	.910	.046	.042	16923663.2	16924825.1	16924472.3
Second Order	226676.7*	398	.915	.907	.047	.046	16944966.3	16945981.7	16945673.4
Bifactor	193636.5*	375	.927	.915	.045	.037	16900725.6	16901981.8	16901600.4

Note: All difference statistics indicate difference from the multifactor model.
CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. SSBIC = Sample Size-Adjusted Bayesian Information Criterion.
*p<.001.

Table 2.4

Fit Statistics for Step 3: Multilevel CFA

	χ^2	df	$\Delta\chi^2$	CFI	ΔCFI	TLI	ΔTLI	RMSEA	ΔRMSEA	SRMR Within	ΔSRMR Within	SRMR Between	ΔSRMR Between	AIC	ΔAIC	BIC	ΔBIC	SSBIC	ΔSSBIC
MLCFA																			
Bifactor	191151.3*	750		.919		.906		.031		.039	.	.067		16737496.6		16739694.8		16739027.4	
Cross-Level Invariance (Loadings Equated)																			
Free	191151.3*	750		.919		.906		.031		.039	.	.067		16737496.6		16739694.8		16739027.4	
Invariant	193624.9*	802	2224.6*	.918	−.001	.911	.005	.030	−.001	.039	.000	.269	.202	16740036.0	2539.4	16741689.8	1995.0	16741187.7	2160.3
Partial	191353.1*	771	77.8*	.919	.000	.908	.002	.031	.000	.039	.000	.068	.001	16737561.6	65.0	16739540.0	−154.8	16738939.3	−88.1
Cluster Bias (Between-Schools Model Residual Variances Fixed to 0)																			
Free	191353.1*	771		.919		.908		.031		.039		.068		16737561.6		16739540.0		16738939.3	
Invariant	422568.2*	801	27612.1*	.820	−.099	.805	−.103	.045	.014	.039	.000	.147	.079	16766793.3	29231.7	16768457.7	28917.7	16767952.3	29013.0
Partial	191430.0*	777	33.9*	.919	.000	.909	.001	.031	.000	.039	.000	.068	.000	16737576.5	14.9	16739492.1	−47.9	16738910.5	−28.8

Note: All difference statistics indicate difference from the Free Model.
CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. SSBIC = Sample Size-Adjusted Bayesian Information Criterion.
*p<.001.

Table 2.5

Final Multilevel Model Results from Step 3 (Partial Invariance) (N = 259,778)

	Within (student level)					Between (school level)				
Relation/Variable	Estimate	SE	Ratio	*p*	Std.	Estimate	SE	Ratio	*p*	Std.
Factor Loadings										
SA by										
Q1	.591	.005	122.5	<.001	.712	.444	.013	35.0	<.001	.949
Q2	*.482*	*.004*	*116.3*	*<.001*	*.558*	*.482*	*.004*	*116.3*	*<.001*	*.934*
Q4	.162	.002	67.9	<.001	.215	.229	.016	14.3	<.001	.584
PR by										
Q6	*.298*	*.003*	*111.9*	*<.001*	*.398*	*.298*	*.003*	*111.9*	*<.001*	*.213*
Q7	*.403*	*.003*	*122.7*	*<.001*	*.504*	*.403*	*.003*	*122.7*	*<.001*	*.363*
Q9	*.424*	*.005*	*88.6*	*<.001*	*.630*	*.424*	*.005*	*88.6*	*<.001*	*.456*
Q20	.150	.004	37.9	<.001	.245	.489	.037	13.2	<.001	.407
Q26	.158	.004	37.9	<.001	.244	.559	.046	12.2	<.001	.419
AR by										
Q11	*.381*	*.003*	*117.6*	*<.001*	*.448*	*.381*	*.003*	*117.6*	*<.001*	*.363*
Q12	*.486*	*.003*	*156.1*	*<.001*	*.550*	*.486*	*.003*	*156.1*	*<.001*	*.393*
Q13	.548	.003	196.8	<.001	.603	.468	.011	42.6	<.001	.383
Q14	.554	.003	181.7	<.001	.590	.496	.014	36.7	<.001	.410
RD by										
Q15	*.523*	*.003*	*160.2*	*<.001*	*.625*	*.523*	*.003*	*160.2*	*<.001*	*.401*
Q16	*.538*	*.003*	*167.7*	*<.001*	*.655*	*.538*	*.003*	*167.7*	*<.001*	*.434*
Q17	*.480*	*.004*	*117.1*	*<.001*	*.528*	*.480*	*.004*	*117.1*	*<.001*	*.357*
Q18	*.374*	*.004*	*88.5*	*<.001*	*.403*	*.374*	*.004*	*88.5*	*<.001*	*.269*
Q19	.419	.004	111.0	<.001	.441	.544	.022	24.4	<.001	.399
PE by										
Q28	.306	.005	62.3	<.001	.363	.717	.091	7.9	<.001	.435
Q29	*.501*	*.005*	*99.7*	*<.001*	*.514*	*.501*	*.005*	*99.7*	*<.001*	*.329*
Q30	*.262*	*.003*	*81.4*	*<.001*	*.330*	*.262*	*.003*	*81.4*	*<.001*	*.221*
Q31	*.308*	*.005*	*59.1*	*<.001*	*.319*	*.308*	*.005*	*59.1*	*<.001*	*.231*
PS by										
Q32_REV	*.262*	*.006*	*46.5*	*<.001*	*.253*	*.262*	*.006*	*46.5*	*<.001*	*.247*
Q36_REV	.462	.005	94.6	<.001	.493	1.381	.196	7.1	<.001	.470
Q37_REV	*.442*	*.006*	*76.8*	*<.001*	*.438*	*.442*	*.006*	*76.8*	*<.001*	*.293*
Q38_REV	.663	.006	104.0	<.001	.560	1.311	.190	6.9	<.001	.497
OS by										
Q87	*.407*	*.005*	*79.3*	*<.001*	*.528*	*.407*	*.005*	*79.3*	*<.001*	*.347*
Q88	*.435*	*.005*	*92.5*	*<.001*	*.572*	*.435*	*.005*	*92.5*	*<.001*	*.442*
Q89	.329	.005	72.0	<.001	.360	.017	.031	0.6	>.05	.013
Q90	*.312*	*.005*	*66.3*	*<.001*	*.327*	*.312*	*.005*	*66.3*	*<.001*	*.262*
Q92	*.280*	*.004*	*65.0*	*<.001*	*.367*	*.280*	*.004*	*65.0*	*<.001*	*.449*
SC by										
Q1	.406	.003	152.4	<.001	.489	.094	.016	5.9	<.001	.301
Q2	.385	.002	157.5	<.001	.446	-.015	.019	-0.8	>.05	-.044
Q4	.363	.003	137.9	<.001	.484	.124	.012	10.5	<.001	.470
Q6	*.339*	*.002*	*138.0*	*<.001*	*.452*	*.339*	*.002*	*138.0*	*<.001*	*.889*
Q7	*.261*	*.003*	*89.7*	*<.001*	*.327*	*.261*	*.003*	*89.7*	*<.001*	*.860*
Q9	*.209*	*.003*	*76.9*	*<.001*	*.310*	*.209*	*.003*	*76.9*	*<.001*	*.820*
Q20	*.299*	*.003*	*97.2*	*<.001*	*.487*	*.299*	*.003*	*97.2*	*<.001*	*.909*
Q26	*.326*	*.003*	*115.7*	*<.001*	*.503*	*.326*	*.003*	*115.7*	*<.001*	*.893*
Q11	.529	.004	144.8	<.001	.623	.500	.008	59.0	<.001	.909
Q12	.576	.003	172.1	<.001	.651	.594	.006	104.5	<.001	.916
Q13	*.591*	*.003*	*194.7*	*<.001*	*.651*	*.591*	*.003*	*194.7*	*<.001*	*.923*
Q14	.614	.003	201.2	<.001	.654	.576	.006	98.7	<.001	.908

Q15	.460	.003	156.2	<.001	.549	.491	.005	90.0	<.001	.914
Q16	*.459*	*.003*	*156.6*	*<.001*	*.559*	*.459*	*.003*	*156.6*	*<.001*	*.899*
Q17	*.501*	*.003*	*169.6*	*<.001*	*.551*	*.501*	*.003*	*169.6*	*<.001*	*.904*
Q18	*.511*	*.003*	*178.7*	*<.001*	*.551*	*.511*	*.003*	*178.7*	*<.001*	*.893*
Q19	.566	.003	193.5	<.001	.595	.499	.010	51.1	<.001	.887
Q28	.505	.005	96.9	<.001	.599	.891	.029	30.5	<.001	.863
Q29	.486	.003	142.8	<.001	.499	.732	.031	23.3	<.001	.767
Q30	.455	.004	102.9	<.001	.574	.684	.020	34.4	<.001	.924
Q31	.516	.003	177.5	<.001	.534	.742	.024	30.7	<.001	.888
Q32_REV	.202	.005	38.1	<.001	.195	.402	.016	24.9	<.001	.826
Q36_REV	.260	.005	47.4	<.001	.278	1.152	.036	32.1	<.001	.855
Q37_REV	.202	.004	47.0	<.001	.200	.501	.026	19.4	<.001	.724
Q38_REV	.286	.006	49.8	<.001	.242	.961	.037	26.2	<.001	.794
Q87	.395	.005	87.7	<.001	.513	.557	.019	30.1	<.001	.901
Q88	.398	.004	105.1	<.001	.523	.464	.017	27.8	<.001	.895
Q89	.478	.004	131.6	<.001	.523	.668	.015	45.5	<.001	.957
Q90	*.501*	*.003*	*159.0*	*<.001*	*.526*	*.501*	*.003*	*159.0*	*<.001*	*.801*
Q92	.316	.003	111.7	<.001	.415	.254	.011	23.2	<.001	.775
Factor Variances/ Covariances										
SA	1.000	—[a]	—	—	1.000	.049	.005	10.2	<.001	1.000
PR	1.000	—[a]	—	—	1.000	.008	.001	5.6	<.001	1.000
AR	1.000	—[a]	—	—	1.000	.030	.003	9.4	<.001	1.000
RD	1.000	—[a]	—	—	1.000	.019	.002	9.3	<.001	1.000
PE	1.000	—[a]	—	—	1.000	.043	.008	5.5	<.001	1.000
PS	1.000	—[a]	—	—	1.000	.023	.007	3.4	.001	1.000
OS	1.000	—[a]	—	—	1.000	.030	.003	8.8	<.001	1.000
SC	1.000	—[a]	—	—	1.000	.110	.007	15.7	<.001	1.000
SC with										
SA	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
AR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
SA with										
PR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
AR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PR with										
AR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
AR with										
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD with										
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000

PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE with										
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS with										
OS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
Error Variances										
Q1	.174	.005	33.5	<.001	.253	.000	—[a]	—	—	.009
Q2	.365	.004	82.3	<.001	.490	.002	.000	7.7	<.001	.125
Q4	.406	.003	147.0	<.001	.720	.003	.000	10.5	<.001	.438
Q6	.358	.004	83.1	<.001	.637	.003	.000	10.7	<.001	.165
Q7	.409	.005	88.2	<.001	.639	.001	.000	8.2	<.001	.129
Q9	.230	.003	71.8	<.001	.508	.001	.000	6.9	<.001	.121
Q11	.296	.003	88.1	<.001	.411	.001	.000	9.3	<.001	.042
Q12	.215	.002	92.9	<.001	.274	.000	.000	3.6	<.001	.006
Q13	.175	.002	82.5	<.001	.212	.000	—[a]	—	—	.002
Q14	.198	.002	102.4	<.001	.225	.000	.000	5.9	<.001	.008
Q15	.215	.002	87.8	<.001	.307	.000	—[a]	—	—	.003
Q16	.174	.002	87.7	<.001	.258	.000	—[a]	—	—	.003
Q17	.345	.003	133.4	<.001	.417	.002	.000	9.9	<.001	.056
Q18	.460	.004	113.0	<.001	.535	.005	.000	10.9	<.001	.131
Q19	.409	.003	129.2	<.001	.452	.002	.000	9.1	<.001	.054
Q20	.264	.004	67.4	<.001	.702	.000	—[a]	—	—	.008
Q26	.288	.003	90.7	<.001	.687	.000	.000	4.1	<.001	.027
Q28	.363	.005	76.0	<.001	.510	.008	.002	3.3	.001	.067
Q29	.462	.005	90.9	<.001	.487	.030	.003	11.2	<.001	.303
Q30	.353	.005	70.1	<.001	.562	.006	.001	8.9	<.001	.098
Q31	.572	.004	153.7	<.001	.613	.012	.001	11.0	<.001	.158
Q32_REV	.959	.007	128.0	<.001	.898	.007	.001	8.9	<.001	.256
Q36_REV	.598	.006	92.4	<.001	.680	.009	.003	3.0	.002	.047
Q37_REV	.782	.014	54.5	<.001	.768	.020	.002	11.7	<.001	.389
Q38_REV	.879	.012	73.1	<.001	.628	.020	.004	5.6	<.001	.124
Q87	.272	.003	81.0	<.001	.458	.003	.000	9.8	<.001	.068
Q88	.231	.003	66.6	<.001	.399	.000	—[a]	—	—	.003
Q89	.499	.007	73.5	<.001	.597	.004	.001	7.2	<.001	.083
Q90	.558	.004	133.6	<.001	.616	.012	.001	11.8	<.001	.290
Q92	.402	.005	85.2	<.001	.693	.002	.000	10.0	<.001	.198
Intercepts										
Q1	n/a	—	—	—	—	2.957	.005	543.5	<.001	28.421
Q2	n/a	—	—	—	—	2.861	.006	482.1	<.001	24.965
Q4	n/a	—	—	—	—	3.293	.005	720.6	<.001	37.793
Q6	n/a	—	—	—	—	3.260	.006	508.8	<.001	25.787
Q7	n/a	—	—	—	—	3.489	.005	670.8	<.001	34.657
Q9	n/a					3.630	.004	847.1	<.001	43.049
Q11	n/a	—	—	—	—	3.247	.009	366.0	<.001	17.820
Q12	n/a	—	—	—	—	3.053	.010	296.1	<.001	14.197
Q13	n/a	—	—	—	—	2.972	.010	292.0	<.001	13.991
Q14	n/a	—	—	—	—	2.944	.010	290.8	<.001	14.002
Q15	n/a					2.593	.010	268.7	<.001	14.557
Q16	n/a	—	—	—	—	2.612	.009	284.3	<.001	15.435
Q17	n/a	—	—	—	—	2.684	.010	266.4	<.001	14.602
Q18	n/a	—	—	—	—	2.944	.009	314.2	<.001	15.525
Q19	n/a					2.623	.010	260.8	<.001	14.079
Q20	n/a	—	—	—	—	3.589	.005	673.3	<.001	32.964
Q26	n/a	—	—	—	—	3.432	.006	582.2	<.001	28.392

Q28	n/a					3.103	.017	186.1	<.001	9.065
Q29	n/a	—	—	—	—	2.628	.015	171.0	<.001	8.315
Q30	n/a	—	—	—	—	3.214	.012	266.5	<.001	13.093
Q31	n/a	—	—	—	—	2.328	.014	170.5	<.001	8.408
Q32_REV	n/a	—	—	—	—	3.115	.008	384.6	<.001	19.334
Q36_REV	n/a	—	—	—	—	2.412	.022	111.3	<.001	5.406
Q37_REV	n/a	—	—	—	—	3.331	.011	293.4	<.001	14.537
Q38_REV	n/a					2.350	.020	120.2	<.001	5.854
Q87	n/a	—	—	—	—	3.379	.010	342.5	<.001	16.503
Q88	n/a	—	—	—	—	3.461	.008	412.8	<.001	20.133
Q89	n/a	—	—	—	—	2.994	.011	262.5	<.001	12.949
Q90	n/a	—	—	—	—	2.837	.009	299.0	<.001	13.686
Q92	n/a	—	—	—	—	3.506	.005	677.3	<.001	32.285

Note. Estimate = unstandardized value; SE = standard error; Ratio = estimate / SE (i.e., Wald statistic); Std = fully standardized value; REV = reversed item. Italicized factor loadings were constrained across levels.
[a]Dashes indicate a parameter fixed for model identification, which was not tested for statistical significance.

Table 2.6

Fit Statistics for Step 4 and Step 5: Within- and Between-Level Invariance

	χ^2	df	$\Delta\chi^2$	CFI	ΔCFI	TLI	ΔTLI	RMSEA	ΔRMSEA	SRMR Within	ΔSRMR Within	SRMR Between	ΔSRMR Between	AIC	ΔAIC	BIC	ΔBIC	SSBIC	ΔSSBIC
Within-Level Invariance (Invariance across Racial/Ethnic Groups)																			
Invariant	199306.3*	865		.916		.904		.030		.035		.091		16728564.9		16730815.4		16730132.2	
Partial	191831.5*	797	10983.6*	.919	.003	.900	−.004	.030	.000	.034	−.001	.076	−.015	16717045.8	−11519.1	16720008.2	−10807.3	16719108.8	−11023.4
Between-Level Invariance (Invariance across School Race/Ethnicity and SES Composition)																			
Invariant	194918.0*	863		.920		.899		.029		.034		.056		16715989.2		16719202.7		16718227.1	
Partial	193043.6*	840	1050.7*	.920	.000	.898	−.001	.030	.001	.034	.000	.042	−.014	16714929.3	−1059.9	16718383.6	−819.1	16717334.8	−892.2

Note: All difference statistics indicate difference from the Free Model.
CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. SSBIC = Sample Size-Adjusted Bayesian Information Criterion.
*$p<.001$.

Table 2.7

Reliability Statistics (including Omega, OmegaH, and H) of the Final Measurement Model from Step 3 (N = 259,778)

	School Climate	School Attachment	Peer Relations	Adult Relations	Respect for Diversity	Physical Environment	Physical Safety	Organizational Structure
Within (student level)								
Omega/OmegaS	.938	.742	.726	.911	.883	.768	.566	.798
OmegaH/OmegaHS	.828	.388	.353	.382	.417	.249	.444	.340
Relative Omega	.882	.523	.486	.419	.472	.324	.784	.426
H	.916	.605	.568	.642	.689	.427	.520	.565
Between (school level)								
Omega/OmegaS	.994	.920	.981	.996	.990	.955	.939	.970
OmegaH/OmegaHS	.959	.847	.150	.152	.145	.106	.170	.106
Relative Omega	.965	.920	.153	.152	.146	.111	.182	.109
H	.990	.943	.466	.415	.456	.316	.435	.414

Note: OmegaS = subscale omega; OmegaH = omega hierarchical; omega HS = omega hierarchical subscale; *H* = measure of construct replicability.

Table 2.8

Parameter Estimates of the Final Structural Model (N = 259,778)

	Within (student level)					Between (school level)				
Relation/Variable	Estimate	SE	Ratio	*p*	Std.	Estimate	SE	Ratio	*p*	Std.
Factor Loadings										
SA by										
Q1	.587	.005	120.039	<.001	.717	.548	.028	19.922	<.001	1.227
Q2	.478	.004	114.720	<.001	.559	.478	.004	114.720	<.001	1.061
Q4	.159	.002	67.740	<.001	.215	.182	.017	10.747	<.001	.493
PR by										
Q6	.297	.003	111.275	<.001	.399	.297	.003	111.275	<.001	.209
Q7	.402	.003	124.392	<.001	.507	.402	.003	124.392	<.001	.354
Q9	.427	.005	90.778	<.001	.638	.427	.005	90.778	<.001	.406
Q20	.146	.004	38.444	<.001	.240	.563	.043	13.060	<.001	.495
Q26	.154	.004	38.501	<.001	.239	.574	.047	12.132	<.001	.463
AR by										
Q11	.377	.003	120.114	<.001	.446	.377	.003	120.114	<.001	.387
Q12	.484	.003	159.342	<.001	.549	.484	.003	159.342	<.001	.404
Q13	.546	.003	198.878	<.001	.604	.464	.013	35.688	<.001	.397
Q14	.552	.003	185.972	<.001	.590	.512	.016	31.841	<.001	.444
RD by										
Q15	.523	.003	159.429	<.001	.628	.523	.003	159.429	<.001	.527
Q16	.537	.003	166.455	<.001	.658	.537	.003	166.455	<.001	.567
Q17	.479	.004	117.176	<.001	.529	.479	.004	117.176	<.001	.435
Q18	.375	.004	90.143	<.001	.405	.375	.004	90.143	<.001	.362
Q19	.419	.004	111.649	<.001	.442	.551	.023	23.886	<.001	.523
PE by										
Q28	.307	.005	62.562	<.001	.368	.682	.061	11.122	<.001	.525
Q29	.500	.005	100.684	<.001	.518	.500	.005	100.684	<.001	.399
Q30	.262	.003	82.454	<.001	.334	.262	.003	82.454	<.001	.289
Q31	.310	.005	59.363	<.001	.324	.310	.005	59.363	<.001	.290
PS by										
Q32_REV	.265	.006	47.506	<.001	.258	.265	.006	47.506	<.001	.096
Q36_REV	.465	.005	96.489	<.001	.498	4.002	1.228	3.258	.001	.529
Q37_REV	.440	.006	77.266	<.001	.437	.440	.006	77.266	<.001	.144
Q38_REV	.658	.006	103.904	<.001	.558	3.763	1.187	3.169	.002	.586
OD by										
Q87	.403	.005	79.158	<.001	.540	.403	.005	79.158	<.001	.359
Q88	.435	.005	92.216	<.001	.591	.435	.005	92.216	<.001	.473
Q89	.328	.005	71.915	<.001	.371	-.008	.037	-.222	.824	-.006
Q90	.309	.005	66.040	<.001	.336	.309	.005	66.040	<.001	.286
Q92	.280	.004	67.199	<.001	.379	.280	.004	67.199	<.001	.466
SC by										
Q1	.407	.003	152.714	<.001	.500	.363	.020	18.493	<.001	.990
Q2	.386	.002	160.040	<.001	.455	.318	.019	16.885	<.001	.861
Q4	.366	.003	139.789	<.001	.496	.287	.014	20.474	<.001	.951
Q6	.338	.002	136.155	<.001	.459	.338	.002	136.155	<.001	.784
Q7	.260	.003	89.899	<.001	.332	.260	.003	89.899	<.001	.757
Q9	.208	.003	76.992	<.001	.315	.208	.003	76.992	<.001	.654
Q20	.298	.003	96.843	<.001	.495	.298	.003	96.843	<.001	.864
Q26	.327	.003	115.609	<.001	.514	.327	.003	115.609	<.001	.871
Q11	.527	.004	144.951	<.001	.632	.475	.010	47.520	<.001	.908
Q12	.574	.003	174.001	<.001	.660	.599	.006	93.322	<.001	.932
Q13	.589	.003	195.544	<.001	.660	.589	.003	195.544	<.001	.939
Q14	.611	.003	200.338	<.001	.662	.591	.008	75.311	<.001	.955

Q15	.457	.003	156.481	<.001	.557	.515	.007	71.465	<.001	.788
Q16	.457	.003	156.171	<.001	.567	.457	.003	156.171	<.001	.732
Q17	.499	.003	169.018	<.001	.559	.499	.003	169.018	<.001	.687
Q18	.511	.003	179.001	<.001	.561	.511	.003	179.001	<.001	.750
Q19	.563	.003	193.882	<.001	.603	.574	.012	48.790	<.001	.827
Q28	.503	.005	97.296	<.001	.608	.848	.046	18.359	<.001	.716
Q29	.485	.003	143.634	<.001	.507	.772	.043	17.982	<.001	.675
Q30	.453	.004	102.924	<.001	.583	.678	.026	26.123	<.001	.819
Q31	.513	.003	179.566	<.001	.541	.857	.034	25.157	<.001	.879
Q32_REV	.202	.005	38.149	<.001	.199	.362	.024	14.759	<.001	.613
Q36_REV	.259	.006	46.777	<.001	.281	.991	.057	17.308	<.001	.615
Q37_REV	.198	.004	47.170	<.001	.200	.199	.025	7.892	<.001	.307
Q38_REV	.283	.006	49.419	<.001	.244	.817	.052	15.655	<.001	.598
Q87	.398	.005	87.886	<.001	.526	.563	.028	20.303	<.001	.779
Q88	.399	.004	105.090	<.001	.534	.492	.024	20.554	<.001	.830
Q89	.477	.004	132.240	<.001	.532	.735	.022	33.115	<.001	.852
Q90	.502	.003	159.588	<.001	.537	.502	.003	159.588	<.001	.720
Q92	.316	.003	112.004	<.001	.423	.276	.017	16.527	<.001	.714
Factor Variances/ Covariances										
SA	1.000	—[a]	—	—	.975	.013	.001	8.898	<.001	.252
PR	1.000	—[a]	—	—	.985	.005	.001	5.203	<.001	.715
AR	1.000	—[a]	—	—	.989	.015	.002	8.478	<.001	.689
RD	1.000	—[a]	—	—	.992	.015	.001	9.935	<.001	.438
PE	1.000	—[a]	—	—	.981	.045	.006	7.301	<.001	.693
PS	1.000	—[a]	—	—	.993	.002	.001	1.604	.109	.603
OD	1.000	—[a]	—	—	.936	.026	.003	7.852	<.001	.787
SC	1.000	—[a]	—	—	.962	.045	.004	12.369	<.001	.580
SC with										
SA	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
AR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
SA with										
PR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
AR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PR with										
AR	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
AR with										
RD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
RD with										
PE	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000

PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PE with										
PS	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
OD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
PS with										
OD	.000	—[a]	—	—	.000	.000	—[a]	—	—	.000
Error Variances										
Q1	.176	.005	33.904	<.001	.255	.000	—[a]	—	—	.010
Q2	.362	.005	79.803	<.001	.483	.001	.000	6.762	<.001	.079
Q4	.403	.003	146.341	<.001	.714	.002	.000	9.316	<.001	.323
Q6	.358	.004	83.379	<.001	.637	.001	.000	8.687	<.001	.096
Q7	.409	.005	88.439	<.001	.640	.001	.000	8.260	<.001	.136
Q9	.228	.003	72.070	<.001	.502	.001	.000	6.456	<.001	.095
Q11	.296	.003	88.358	<.001	.409	.001	.000	9.593	<.001	.066
Q12	.215	.002	92.925	<.001	.274	.000	.000	3.998	<.001	.010
Q13	.175	.002	82.553	<.001	.211	.000	—[a]	—	—	.003
Q14	.198	.002	102.388	<.001	.224	.000	.000	4.631	<.001	.009
Q15	.215	.002	88.052	<.001	.307	.000	—[a]	—	—	.003
Q16	.174	.002	87.948	<.001	.258	.000	—[a]	—	—	.003
Q17	.345	.003	133.457	<.001	.417	.002	.000	8.960	<.001	.038
Q18	.458	.004	112.167	<.001	.532	.004	.000	11.128	<.001	.109
Q19	.408	.003	128.959	<.001	.451	.001	.000	7.819	<.001	.037
Q20	.263	.004	67.550	<.001	.699	.000	—[a]	—	—	.011
Q26	.287	.003	90.880	<.001	.683	.000	.000	3.900	<.001	.030
Q28	.362	.005	75.924	<.001	.509	.008	.002	3.624	<.001	.072
Q29	.463	.005	91.432	<.001	.487	.029	.003	11.428	<.001	.287
Q30	.353	.005	70.315	<.001	.561	.005	.001	9.084	<.001	.102
Q31	.571	.004	154.104	<.001	.611	.010	.001	9.962	<.001	.138
Q32_REV	.957	.007	128.078	<.001	.895	.006	.001	8.444	<.001	.231
Q36_REV	.595	.006	92.463	<.001	.677	.009	.003	2.713	.007	.046
Q37_REV	.774	.014	54.930	<.001	.757	.006	.001	10.645	<.001	.195
Q38_REV	.882	.012	73.638	<.001	.630	.017	.003	5.607	<.001	.118
Q87	.273	.003	80.965	<.001	.459	.002	.000	9.466	<.001	.061
Q88	.230	.003	66.603	<.001	.398	.000	—[a]	—	—	.004
Q89	.498	.007	73.414	<.001	.596	.005	.001	7.277	<.001	.080
Q90	.557	.004	132.817	<.001	.613	.010	.001	10.877	<.001	.264
Q92	.402	.005	85.137	<.001	.692	.002	.000	9.913	<.001	.201
Intercepts										
Q1	n/a	—	—	—	—	2.918	.019	150.051	<.001	28.521
Q2	n/a	—	—	—	—	2.731	.018	155.216	<.001	26.494
Q4	n/a	—	—	—	—	3.353	.015	226.791	<.001	39.711
Q6	n/a	—	—	—	—	3.487	.015	232.491	<.001	28.999
Q7	n/a	—	—	—	—	3.671	.011	340.148	<.001	38.195
Q9	n/a					3.782	.009	437.148	<.001	42.562
Q11	n/a	—	—	—	—	3.466	.022	154.111	<.001	23.709
Q12	n/a	—	—	—	—	3.303	.028	116.749	<.001	18.418
Q13	n/a	—	—	—	—	3.226	.028	115.134	<.001	18.415
Q14	n/a	—	—	—	—	3.168	.029	109.978	<.001	18.325
Q15	n/a					2.909	.027	107.392	<.001	15.936
Q16	n/a	—	—	—	—	2.911	.025	115.926	<.001	16.705
Q17	n/a	—	—	—	—	3.030	.026	116.850	<.001	14.952
Q18	n/a	—	—	—	—	3.287	.026	126.545	<.001	17.288
Q19	n/a					2.927	.030	99.111	<.001	15.103
Q20	n/a	—	—	—	—	3.778	.012	313.524	<.001	39.224
Q26	n/a	—	—	—	—	3.638	.013	273.155	<.001	34.699

Q28	n/a					3.491	.044	79.196	<.001	10.548
Q29	n/a	—	—	—	—	2.946	.043	68.142	<.001	9.210
Q30	n/a	—	—	—	—	3.490	.031	113.401	<.001	15.092
Q31	n/a	—	—	—	—	2.683	.039	69.282	<.001	9.859
Q32_REV	n/a	—	—	—	—	3.421	.021	165.232	<.001	20.761
Q36_REV	n/a	—	—	—	—	3.278	.059	55.638	<.001	7.286
Q37_REV	n/a	—	—	—	—	3.620	.021	170.391	<.001	20.007
Q38_REV	n/a					3.039	.057	53.146	<.001	7.957
Q87	n/a	—	—	—	—	3.668	.029	127.089	<.001	18.160
Q88	n/a	—	—	—	—	3.655	.026	138.245	<.001	22.069
Q89	n/a	—	—	—	—	3.329	.032	102.826	<.001	13.815
Q90	n/a	—	—	—	—	3.013	.029	103.721	<.001	15.463
Q92	n/a	—	—	—	—	3.612	.016	220.464	<.001	33.410
Latent Factor Direct Effects										
SA on										
AA	.272	.017	15.881	<.001	.127	n/a	—	—	—	—
Lat	-.189	.023	-8.080	<.001	−.064	n/a	—	—	—	—
As	.222	.027	8.086	<.001	.047	n/a	—	—	—	—
Oth	.125	.025	4.932	<.001	.027	n/a	—	—	—	—
HH	n/a	—	—	—	—	.119	.050	2.385	.017	.095
Prop Minority	n/a	—	—	—	—	.177	.045	3.951	<.001	.219
Prop Low SES	n/a	—	—	—	—	.924	.066	14.044	<.001	.747
PR on										
AA	-.047	.020	-2.319	.020	−.022	n/a	—	—	—	—
Lat	-.369	.028	-13.202	<.001	−.126	n/a	—	—	—	—
As	-.130	.029	-4.454	<.001	−.027	n/a	—	—	—	—
Oth	-.163	.031	-5.337	<.001	−.036	n/a	—	—	—	—
HH	n/a	—	—	—	—	.133	.036	3.748	<.001	.288
Prop Minority	n/a	—	—	—	—	.105	.027	3.858	<.001	.351
Prop Low SES	n/a	—	—	—	—	-.194	.041	-4.736	<.001	-.423
AR on										
AA	-.076	.026	-2.952	.003	−.035	n/a	—	—	—	—
Lat	-.316	.030	-10.446	<.001	−.108	n/a	—	—	—	—
As	.038	.039	.976	.329	.008	n/a	—	—	—	—
Oth	-.085	.035	-2.399	.016	−.019	n/a	—	—	—	—
HH	n/a	—	—	—	—	.369	.069	5.309	<.001	.450
Prop Minority	n/a	—	—	—	—	-.261	.053	-4.959	<.001	-.492
Prop Low SES	n/a	—	—	—	—	.601	.066	9.069	<.001	.742
RD on										
AA	.140	.017	8.076	<.001	.066	n/a	—	—	—	—
Lat	-.138	.020	-6.942	<.001	−.047	n/a	—	—	—	—
As	.041	.029	1.418	.156	.009	n/a	—	—	—	—
Oth	.013	.027	.492	.623	.003	n/a	—	—	—	—
HH	n/a	—	—	—	—	.108	.053	2.058	.040	.107
Prop Minority	n/a	—	—	—	—	-.037	.043	-.860	.390	-.057
Prop Low SES	n/a	—	—	—	—	-.658	.071	-9.304	<.001	-.662
PE on										
AA	.231	.027	8.588	<.001	.108	n/a	—	—	—	—
Lat	-.177	.032	-5.564	<.001	−.060	n/a	—	—	—	—
As	.080	.047	1.704	.088	.017	n/a	—	—	—	—
Oth	.164	.037	4.494	<.001	.036	n/a	—	—	—	—
HH	n/a	—	—	—	—	.598	.130	4.593	<.001	.429
Prop Minority	n/a	—	—	—	—	-.237	.085	-2.777	.005	-.263
Prop Low SES	n/a	—	—	—	—	-.050	.112	-.449	.654	-.037
PS on										

AA	.151	.018	8.401	<.001	.071	n/a	—	—	—	—
Lat	-.097	.023	-4.167	<.001	−.033	n/a	—	—	—	—
As	.092	.027	3.393	.001	.019	n/a	—	—	—	—
Oth	.050	.025	2.000	.045	.011	n/a	—	—	—	—
HH	n/a	—	—	—	—	-.006	.017	-.341	.733	-.018
Prop Minority	n/a	—	—	—	—	.033	.019	1.701	.089	.155
Prop Low SES	n/a	—	—	—	—	-.232	.079	-2.937	.003	-.721
OD on										
AA	.268	.024	11.301	<.001	.122	n/a	—	—	—	—
Lat	-.558	.032	-17.644	<.001	−.186	n/a	—	—	—	—
As	-.209	.039	-5.420	<.001	−.043	n/a	—	—	—	—
Oth	-.015	.036	-.405	.686	−.003	n/a	—	—	—	—
HH	n/a	—	—	—	—	.425	.071	5.953	<.001	.432
Prop Minority	n/a	—	—	—	—	-.173	.056	-3.108	.002	-.272
Prop Low SES	n/a	—	—	—	—	.127	.079	1.613	.107	.130
SC on										
AA	-.170	.022	-7.732	<.001	−.079	n/a	—	—	—	—
Lat	.447	.030	14.766	<.001	.151	n/a	—	—	—	—
As	.112	.035	3.178	.001	.023	n/a	—	—	—	—
Oth	-.162	.032	-5.058	<.001	−.035	n/a	—	—	—	—
HH	n/a	—	—	—	—	-.251	.079	-3.172	.002	-.164
Prop Minority	n/a	—	—	—	—	-.455	.054	-8.399	<.001	-.461
Prop Low SES	n/a	—	—	—	—	-.425	.090	-4.713	<.001	-.282
Observed Variables										
Direct Effects										
Q1 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	-.231	.028	-8.131	<.001	-.417
Q2 on										
AA	.126	.005	25.981	<.001	.069	n/a	—	—	—	—
LAT	.116	.006	20.596	<.001	.046	n/a	—	—	—	—
AS	.090	.008	11.634	<.001	.022	n/a	—	—	—	—
OTH	.078	.007	11.541	<.001	.020	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	-.061	.013	-4.566	<.001	−.169
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q4 on										
AA	.071	.008	9.358	<.001	.044	n/a	—	—	—	—
LAT	-.219	.011	-20.750	<.001	−.100	n/a	—	—	—	—
AS	-.039	.011	-3.482	<.001	−.011	n/a	—	—	—	—
OTH	-.042	.011	-3.948	<.001	−.013	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q6 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.074	.013	5.670	<.001	.174

Prop Low SES	n/a	—	—	—	—	-.233	.020	-11.953	<.001	−.359
Q7 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	-.054	.010	-5.328	<.001	−.159
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q9 on										
AA	.043	.005	8.579	<.001	.030	n/a	—	—	—	—
LAT	.056	.007	7.919	<.001	.029	n/a	—	—	—	—
AS	-.006	.007	-.742	.458	−.002	n/a	—	—	—	—
OTH	.020	.008	2.455	.014	.007	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	-.093	.009	-10.705	<.001	−.298
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q11 on										
AA	-.029	.005	-6.049	<.001	−.016	n/a	—	—	—	—
LAT	-.056	.006	-9.693	<.001	−.023	n/a	—	—	—	—
AS	.032	.007	4.489	<.001	.008	n/a	—	—	—	—
OTH	-.028	.007	-4.362	<.001	−.007	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q12 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q13 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q14 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.035	.007	5.032	<.001	.058
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q15 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000

Prop Low SES	n/a	—	—	—	—	.142	.019	7.544	<.001	.144
Q16 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.105	.019	5.671	<.001	.112
Q17 on										
AA	.061	.005	12.998	<.001	.032	n/a	—	—	—	—
LAT	.045	.005	9.031	<.001	.017	n/a	—	—	—	—
AS	.006	.009	.734	.463	.001	n/a	—	—	—	—
OTH	.023	.007	3.251	.001	.006	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	-.046	.011	-4.078	<.001	−.064
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q18 on										
AA	.050	.007	6.739	<.001	.025	n/a	—	—	—	—
LAT	-.115	.008	-14.744	<.001	-.043	n/a	—	—	—	—
AS	-.197	.010	-19.167	<.001	-.045	n/a	—	—	—	—
OTH	-.027	.010	-2.545	.011	-.006	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q19 on										
AA	.058	.007	8.866	<.001	.029	n/a	—	—	—	—
LAT	-.001	.006	-.111	.912	.000	n/a	—	—	—	—
AS	-.002	.010	-.159	.874	.000	n/a	—	—	—	—
OTH	.016	.009	1.647	.100	.004	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.237	.025	9.532	<.001	.226
Q20 on										
AA	-.069	.005	-13.168	<.001	−.053	n/a	—	—	—	—
LAT	-.141	.007	-19.847	<.001	−.079	n/a	—	—	—	—
AS	-.093	.009	-10.785	<.001	−.032	n/a	—	—	—	—
OTH	-.043	.008	-5.683	<.001	−.015	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q26 on										
AA	-.053	.006	-9.585	<.001	−.039	n/a	—	—	—	—
LAT	-.196	.009	-22.659	<.001	−.104	n/a	—	—	—	—
AS	-.069	.009	-7.407	<.001	−.023	n/a	—	—	—	—
OTH	-.043	.008	-5.072	<.001	−.015	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q28 on										
AA	-.045	.006	-7.787	<.001	−.025	n/a	—	—	—	—
LAT	-.122	.007	-16.934	<.001	−.050	n/a	—	—	—	—
AS	-.072	.008	-8.986	<.001	−.018	n/a	—	—	—	—
OTH	-.062	.009	-7.243	<.001	−.016	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000

Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q29 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q30 on										
AA	-.054	.006	-8.764	<.001	−.032	n/a	—	—	—	—
LAT	-.123	.009	-14.034	<.001	−.053	n/a	—	—	—	—
AS	-.028	.010	-2.874	.004	−.007	n/a	—	—	—	—
OTH	-.044	.009	-5.102	<.001	−.012	n/a	—	—	—	—
HH	n/a	—	—	—	—	.165	.030	5.510	<.001	.130
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q31 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.132	.037	3.588	<.001	.090
Q32_REV on										
AA	.024	.007	3.412	.001	.011	n/a	—	—	—	—
LAT	-.086	.009	-9.920	<.001	−.029	n/a	—	—	—	—
AS	-.028	.010	-2.816	.005	−.006	n/a	—	—	—	—
OTH	-.046	.011	-4.155	<.001	−.010	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	-.298	.036	-8.295	<.001	−.335
Q36_REV on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	-.173	.042	-4.133	<.001	−.109
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q37_REV on										
AA	-.297	.011	-26.806	<.001	−.139	n/a	—	—	—	—
LAT	-.074	.014	-5.144	<.001	−.025	n/a	—	—	—	—
AS	.020	.013	1.559	.119	.004	n/a	—	—	—	—
OTH	-.176	.013	-13.458	<.001	−.039	n/a	—	—	—	—
HH	n/a	—	—	—	—	.147	.029	5.167	<.001	.149
Prop Minority	n/a	—	—	—	—	.241	.025	9.617	<.001	.377
Prop Low SES	n/a	—	—	—	—	-.734	.047	-15.775	<.001	−.751
Q38_REV on										
AA	-.128	.012	-11.090	<.001	−.051	n/a	—	—	—	—
LAT	.028	.013	2.178	.029	.008	n/a	—	—	—	—
AS	.119	.016	7.253	<.001	.021	n/a	—	—	—	—
OTH	-.039	.015	-2.539	.011	−.007	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000

Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q87 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	-.147	.019	-7.793	<.001	−.134
Q88 on										
AA	-.026	.004	-6.287	<.001	−.016	n/a	—	—	—	—
LAT	.069	.006	11.513	<.001	.031	n/a	—	—	—	—
AS	.057	.007	8.397	<.001	.016	n/a	—	—	—	—
OTH	.014	.007	1.931	.054	.004	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q89 on										
AA	.051	.006	8.262	<.001	.026	n/a	—	—	—	—
LAT	.047	.007	6.947	<.001	.018	n/a	—	—	—	—
AS	.110	.010	11.490	<.001	.026	n/a	—	—	—	—
OTH	.035	.009	4.041	<.001	.009	n/a	—	—	—	—
HH	n/a	—	—	—	—	.204	.039	5.299	<.001	.155
Prop Minority	n/a	—	—	—	—	-.118	.029	-4.058	<.001	−.139
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Q90 on										
AA	.113	.007	16.611	<.001	.056	n/a	—	—	—	—
LAT	.062	.008	7.748	<.001	.022	n/a	—	—	—	—
AS	.094	.013	7.244	<.001	.021	n/a	—	—	—	—
OTH	.070	.010	6.921	<.001	.016	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	-.168	.028	-5.953	<.001	−.244
Prop Low SES	n/a	—	—	—	—	.258	.035	7.296	<.001	.245
Q92 on										
AA	.000	—[a]	—	—	.000	n/a	—	—	—	—
LAT	.000	—[a]	—	—	.000	n/a	—	—	—	—
AS	.000	—[a]	—	—	.000	n/a	—	—	—	—
OTH	.000	—[a]	—	—	.000	n/a	—	—	—	—
HH	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Minority	n/a	—	—	—	—	.000	—[a]	—	—	.000
Prop Low SES	n/a	—	—	—	—	.000	—[a]	—	—	.000
Latent Factor R^2										
SA	.025	.002	10.776	<.001	n/a	.748	.035	21.538	<.001	n/a
PR	.015	.002	6.990	<.001	n/a	.285	.062	4.636	<.001	n/a
AR	.011	.002	5.460	<.001	n/a	.311	.053	5.900	<.001	n/a
RD	.008	.001	6.525	<.001	n/a	.562	.056	9.960	<.001	n/a
PE	.019	.003	6.693	<.001	n/a	.307	.074	4.172	<.001	n/a
PS	.007	.001	5.794	<.001	n/a	.397	.050	8.014	<.001	n/a
OD	.064	.005	13.195	<.001	n/a	.213	.056	3.820	<.001	n/a
SC	.038	.003	10.991	<.001	n/a	.420	.045	9.275	<.001	n/a
Observed Indicator R^2										
Q1	.745	.007	100.921	<.001	n/a	.990	.001	1173.535	<.001	n/a
Q2	.517	.005	98.059	<.001	n/a	.921	.012	74.404	<.001	n/a
Q4	.286	.003	106.705	<.001	n/a	.677	.035	19.614	<.001	n/a
Q6	.363	.003	112.728	<.001	n/a	.904	.012	75.658	<.001	n/a
Q7	.360	.004	100.416	<.001	n/a	.864	.018	48.651	<.001	n/a

Q9	.498	.006	76.578	<.001	n/a	.905	.015	60.008	<.001	n/a
Q11	.591	.002	241.030	<.001	n/a	.934	.008	119.736	<.001	n/a
Q12	.726	.002	382.606	<.001	n/a	.990	.003	374.383	<.001	n/a
Q13	.789	.002	425.816	<.001	n/a	.997	.000	4777.310	<.001	n/a
Q14	.776	.002	452.578	<.001	n/a	.991	.002	476.758	<.001	n/a
Q15	.693	.003	274.905	<.001	n/a	.997	.000	4769.366	<.001	n/a
Q16	.742	.002	304.375	<.001	n/a	.997	.000	4293.999	<.001	n/a
Q17	.583	.003	201.962	<.001	n/a	.962	.005	199.986	<.001	n/a
Q18	.468	.003	176.928	<.001	n/a	.891	.011	80.062	<.001	n/a
Q19	.549	.003	214.931	<.001	n/a	.963	.005	176.818	<.001	n/a
Q20	.301	.003	91.228	<.001	n/a	.989	.001	1362.449	<.001	n/a
Q26	.317	.003	95.967	<.001	n/a	.970	.008	122.566	<.001	n/a
Q28	.491	.004	136.756	<.001	n/a	.928	.020	46.212	<.001	n/a
Q29	.513	.005	112.077	<.001	n/a	.713	.026	27.817	<.001	n/a
Q30	.439	.003	150.605	<.001	n/a	.898	.012	76.884	<.001	n/a
Q31	.389	.003	128.111	<.001	n/a	.862	.016	53.483	<.001	n/a
Q32_REV	.105	.002	42.541	<.001	n/a	.769	.027	28.899	<.001	n/a
Q36_REV	.323	.005	59.516	<.001	n/a	.954	.017	55.027	<.001	n/a
Q37_REV	.243	.004	54.820	<.001	n/a	.805	.020	39.889	<.001	n/a
Q38_REV	.370	.007	52.837	<.001	n/a	.882	.022	39.969	<.001	n/a
Q87	.541	.004	132.165	<.001	n/a	.939	.008	123.845	<.001	n/a
Q88	.602	.004	167.723	<.001	n/a	.996	.000	3085.961	<.001	n/a
Q89	.404	.004	114.124	<.001	n/a	.920	.012	75.916	<.001	n/a
Q90	.387	.003	129.752	<.001	n/a	.736	.024	30.965	<.001	n/a
Q92	.308	.004	85.224	<.001	n/a	.799	.021	38.786	<.001	n/a

Note. Estimate = unstandardized value; SE = standard error; Ratio = estimate / SE (i.e., Wald statistic); Std = fully standardized value; REV = reversed item.

[a]Dashes indicate a parameter fixed for model identification, which was not tested for statistical significance.

Table 2.9

Completely Standardized Factor Loadings, Direct Effects, and Intraclass Correlations of the Final Structural Model (N = 259,778)

			Within (student level)												Between (school level)									
Factor/Variable	Item	SC	SA	PR	AR	RD	PE	PS	OD	AA	Lat	As	Oth	SC	SA	PR	AR	RD	PE	PS	OD	HH	PM	PL
School Climate										**−.08**	**.15**	***.02***	**−.04**									***−.16***	**−.46**	**−.28**
School Attachment										*	*	*	*									***ns***	**.22**	**.75**
	Q1	.50	.72							—	—	—	—	.99	1.23							—	—	−.42
	Q2	.46	.56							.07	.05	.02	.02	.86	1.06							—	−.17	—
	Q4	.50	.22							.04	−.10	*−.01*	−.01	.95	.49							—	—	—
Peer Relations										*	*	*	*									*	*	*
	Q6	.46		.40						—	—	—	—	.78		.21						—	.17	−.36
	Q7	.33		.51						—	—	—	—	.76		.35						—	−.16	—
	Q9	.32		.64						.03	.03	***ns***	***ns***	.65		.41						—	−.30	—
	Q20	.50		.24						−.05	−.08	−.03	−.02	.86		.50						—	—	—
	Q26	.51		.24						−.04	−.10	−.02	−.02	.87		.46						—	—	—
Adult Relations										*	*	*	*									*	*	*
	Q11	.63			.45					−.02	−.02	.01	−.01	.91			.39					—	—	—
	Q12	.66			.55					—	—	—	—	.93			.40					—	—	—
	Q13	.66			.60					—	—	—	—	.94			.40					—	—	—
	Q14	.66			.59					—	—	—	—	.96			.44					—	.06	—
Respect for Diversity										*	*	*	*									*	*	*
	Q15	.56				.63				—	—	—	—	.79				.53				—	—	.14
	Q16	.57				.66				—	—	—	—	.73				.57				—	—	.11
	Q17	.56				.53				.03	.02	***ns***	*.01*	.69				.44				—	−.06	—
	Q18	.56				.41				.03	−.04	−.05	***ns***	.75				.36				—	—	—
	Q19	.60				.44				.03	***ns***	***ns***	***ns***	.83				.52				—	—	.23
Physical Environment										*	*	*	*									*	*	*
	Q28	.61					.37			−.03	−.05	−.02	−.02	.72					.53			—	—	—
	Q29	.51					.52			—	—	—	—	.68					.40			—	—	—
	Q30	.58					.33			−.03	−.05	*−.01*	−.01	.82					.29			.13	—	—
	Q31	.54					.32			—	—	—	—	.88					.29			—	—	.09
Physical Safety										*	*	*	*									*	*	*
	Q32R	.20						.26		*.01*	−.03	*−.01*	−.01	.61						.10		—	—	−.34
	Q36R	.28						.50		—	—	—	—	.62						.53		—	−.11	—
	Q37R	.20						.44		−.14	−.03	***ns***	−.04	.31						.14		.15	.38	−.75
	Q38R	.24						.56		−.05	***ns***	.02	***ns***	.60						.59		—	—	—
Organizational Structure		.								*	*	*	*									*	*	*
	Q87	.53							.54	—	—	—	—	.78							.36	—	—	−.13
	Q88	.53							.59	−.02	.03	.02	***ns***	.83							.47	—	—	—
	Q89	.53							.37	.03	.02	.03	.01	.85							-.01	.16	−.14	—
	Q90	.54							.34	.06	.02	.02	.02	.72							.29	—	−.24	.25
	Q92	.42							.38	—	—	—	—	.71							.47	—	—	—

*As recommended in DeMars (2013), structural estimates for specific factors were not included in this table to avoid misinterpretation, given their low reliability.

Note: Direct effects labeled “*ns*” were not significant at the .01 level (i.e., $p > .01$). Direct effects in italics were significant at the .01 level (i.e., $p < .01$). All other direct effects and all factor loadings were significant at the .001 level (i.e., $p < .001$).

Figure 2.1

A Conceptual Model of School Climate based on Bronfenbrenner's PPCT Model

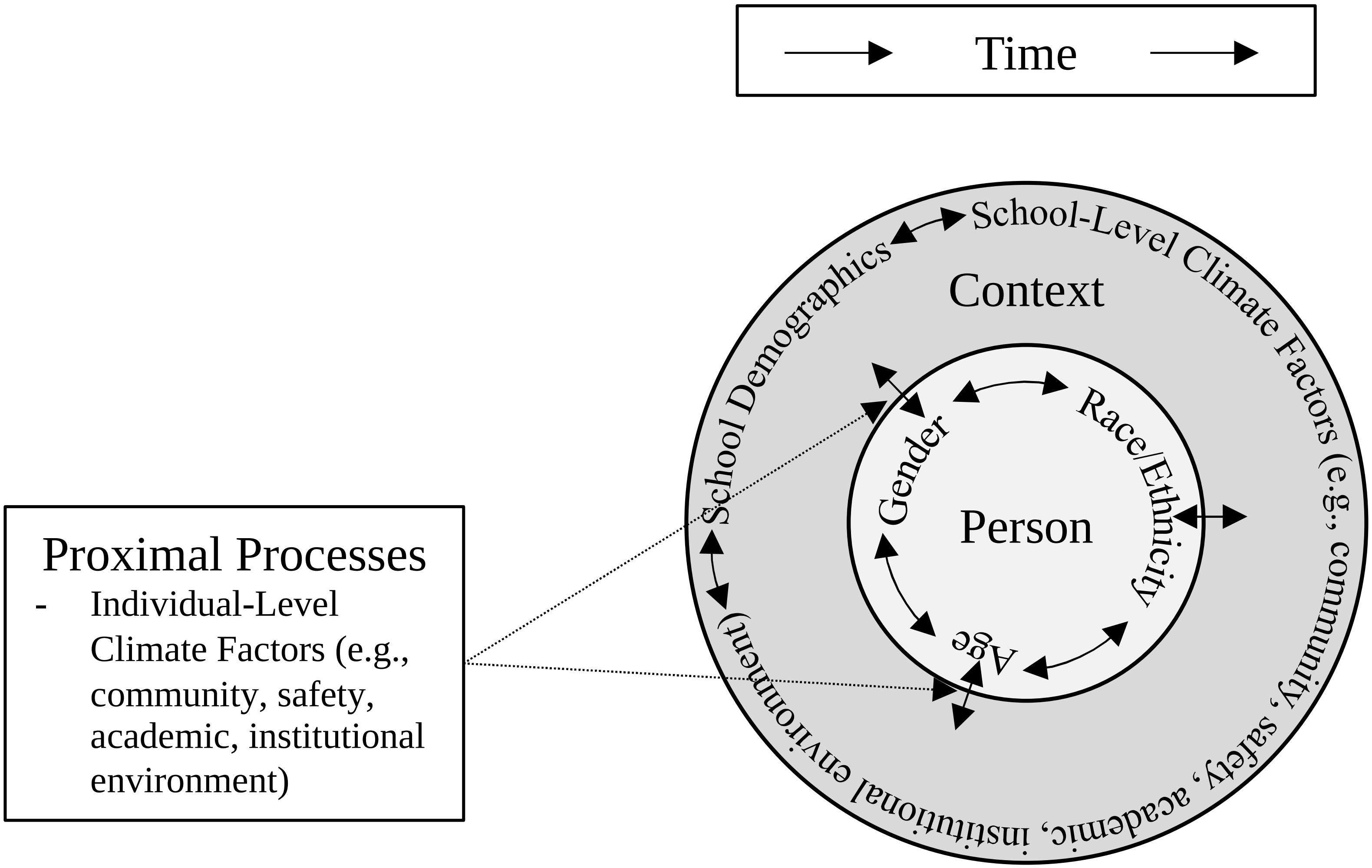


Figure 2.2

A Diagram of the Bifactor GSCS Model from Step 1

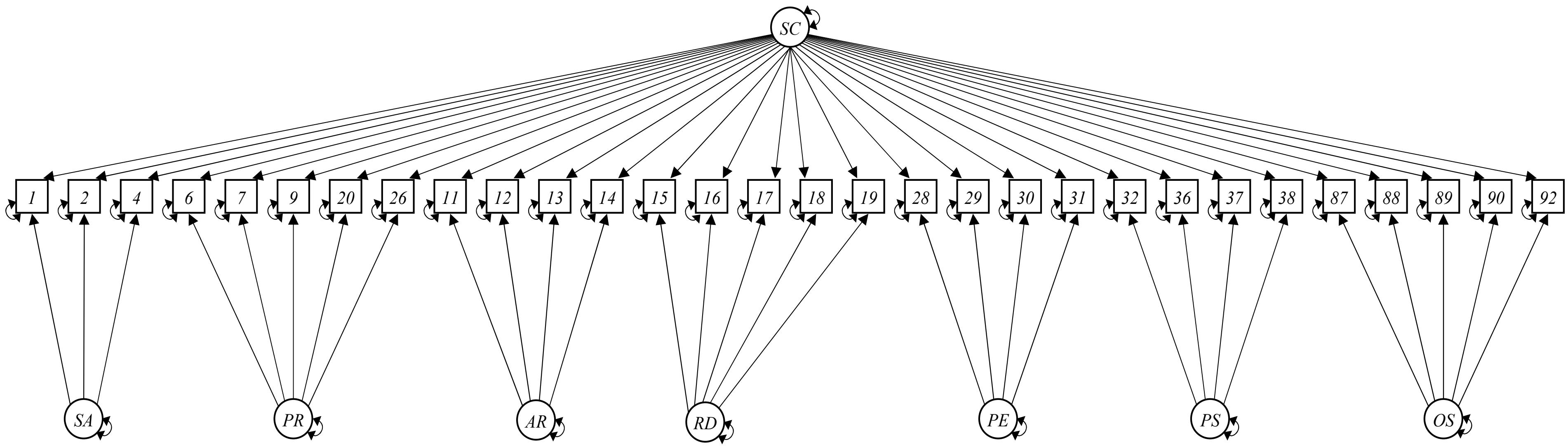


Figure 2.3

A Diagram of the Partially Invariant Multilevel GSCS Model from Step 3

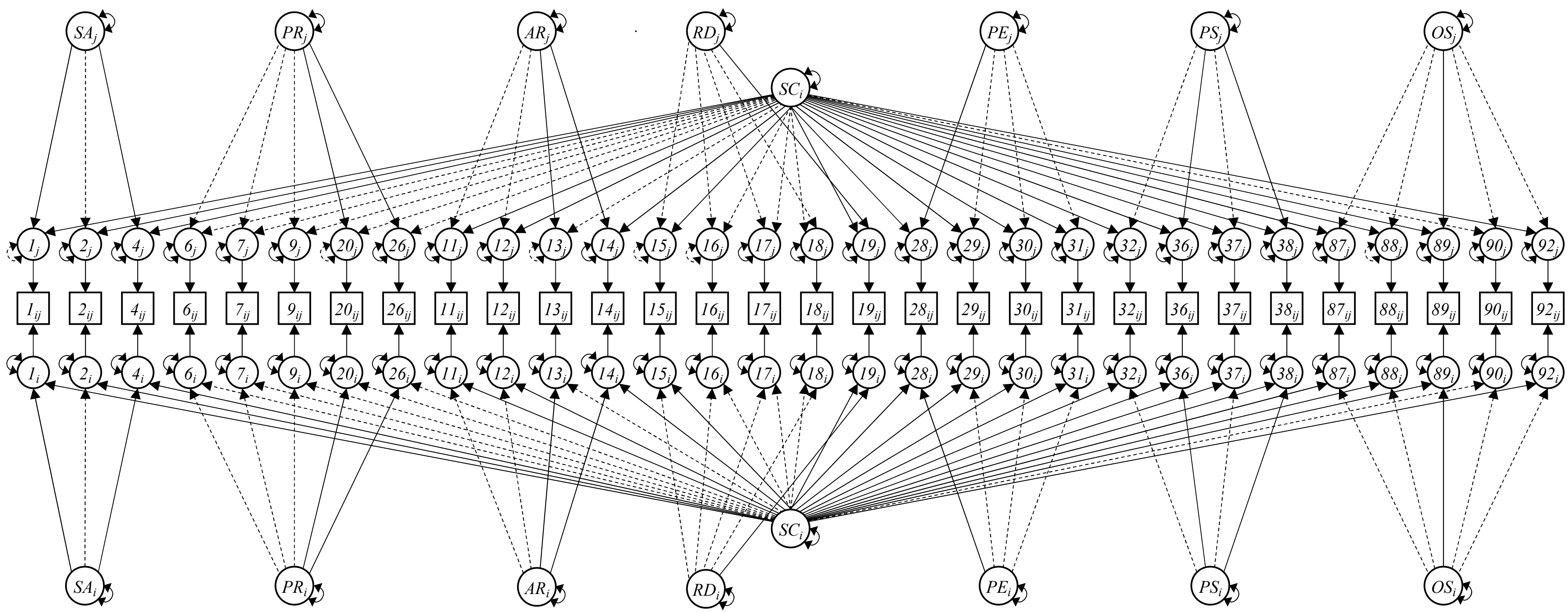


Figure 2.4
Final Model from Step 6

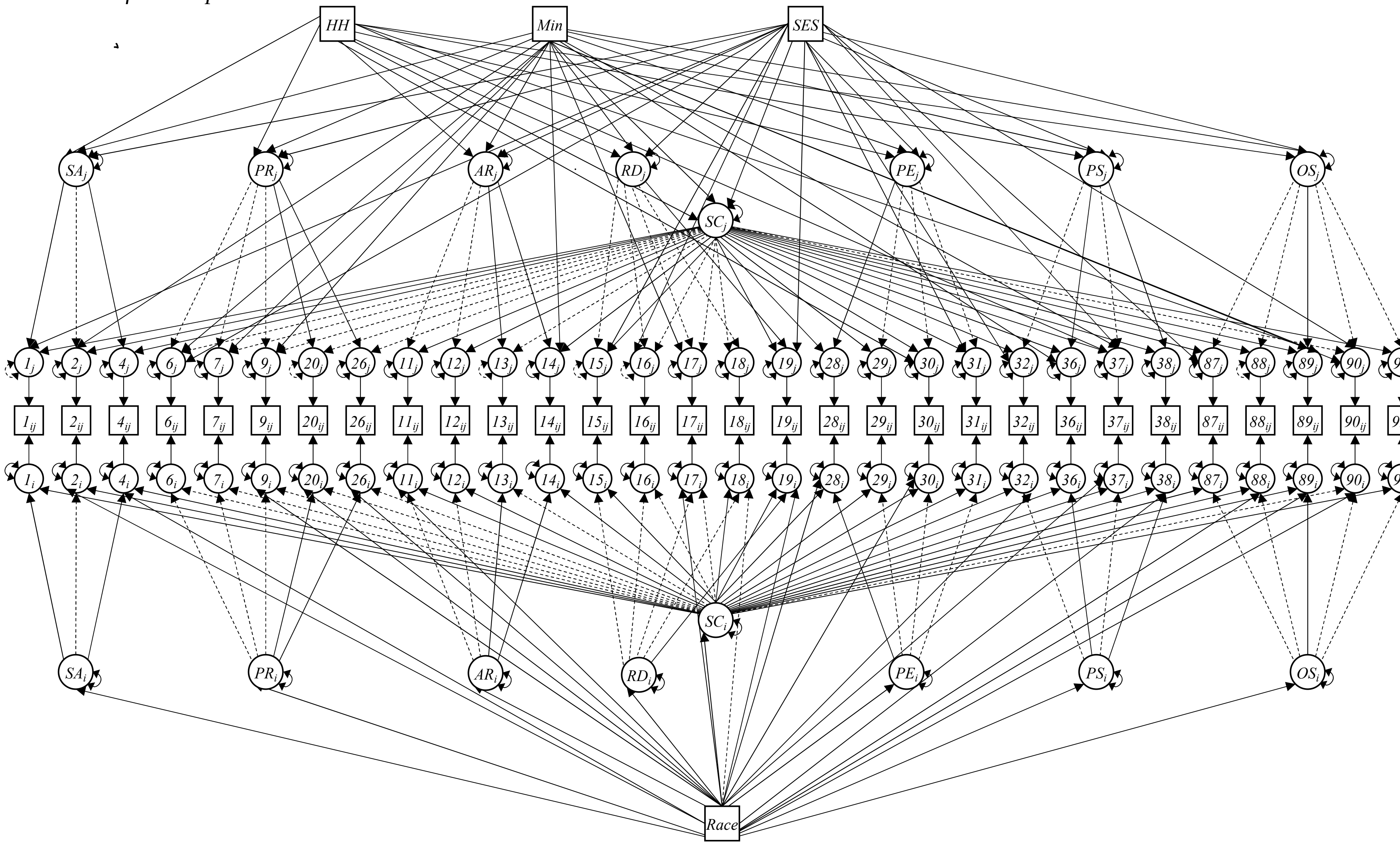
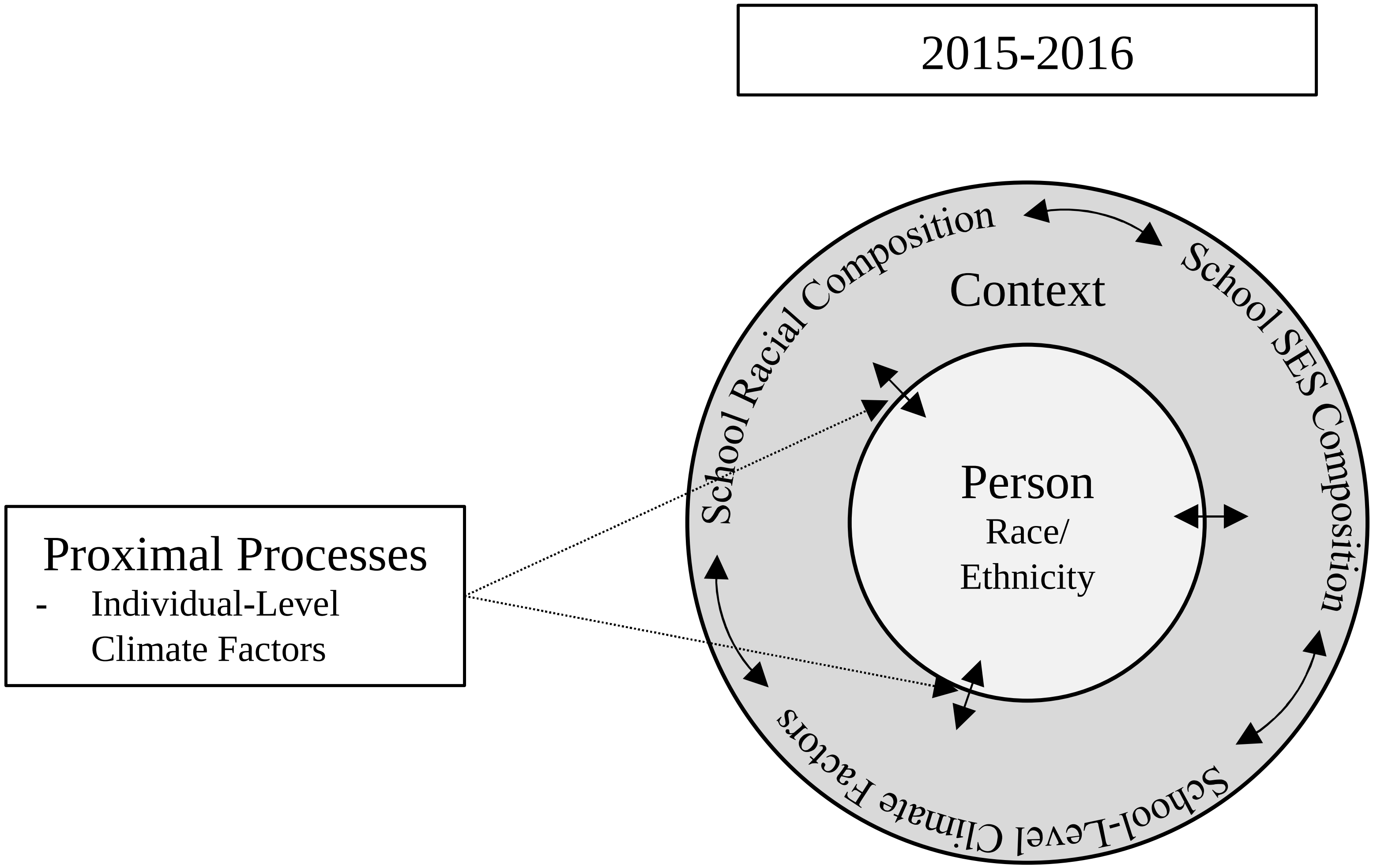


Figure 2.5

A Conceptual Model of the Present Study based on Bronfenbrenner's PPCT Model



APPENDICES

Appendix A

Confirmation of Initial Results using WLSMV Estimation

Step 1: Conventional factor analysis. Conventional CFA results suggested that the theoretical GSCS factor structure of the GSCS yielded better fit than the original GSCS factor structure (e.g., ΔCFI = .029, ΔTLI = .031, and ΔRMSEA = −.008, see Table A1). Model fit statistics of the final model provided support for the validity of the 30-item, 7-factor model (RMSEA = .06, CFI = .96, TLI = .96, and $\chi 2$ (384, N = 259,778) = 354767.5, p < 0.01).

Next, alternative models were specified to determine the best-fitting underlying factor structure. Results are shown in Table A1. A one-factor model, the most parsimonious of the models, yielded poor fit statistics (e.g., CFI = .80, RMSEA = .13). A second order model was specified in which the latent school climate subfactors served as indicators of an overall school climate factor. The DIFFTEST $\chi 2$ procedure in Mplus indicated that the second-order structure did not significantly improve fit over multifactor structure and alternative fit-statistics suggested comparable fit (e.g., ΔCFI = −.001, ΔTLI = .001, and ΔRMSEA = −.001). Next, a bifactor model with one general factor (i.e., School Climate) and seven specific factors was estimated. Following the logic for model comparison from Brunner et al. (2012), the fit of the bifactor model was a significant improvement over the second-order model, DIFFTEST $\chi 2$(23, N = 259,778) = 43397.24, p < .001, and each of the alternative fit indices for this model suggested improved fit over the multifactor model (e.g., ΔRMSEA = –.004, ΔCFI = .005, ΔTLI = .005).

The bifactor model provides the ability to investigate how item variance is partitioned between general and specific factors (Rodriguez et al., 2016). Using WLSMV estimation, the GSCS items were generally good indicators of the general school climate factor, with standardized loadings ranging from .28 (item 32) to .72 (item 13) with only two loadings <.30. When comparing those loadings with the item loadings for seven specific factors, slightly more total variance could be attributed to the general factor than to the specific factors (*M* = .56 & .50, respectively). However, this pattern varied by factor. For example, the adult relations, physical environment, and organizational structure items all loaded more strongly on the general School Climate factor than on their respective specific factors, whereas the physical safety items all loaded more strongly on the specific Physical Safety factor than on the general School Climate factor. Item loadings ranged from .26 to .77 for School Attachment (*M* = .55), from .39 to .61 for Peer Relations (*M* = .49), from .50 to .60 for Adult Relations (*M* = .56), from .46 to .66 for Respect for Diversity (*M* = .55), from .29 to .53 for Physical Environment (*M* = .38), from .29 to .66 for Physical Safety (*M* = .52), and from .39 to .57 for Organizational Structure (*M* = .46).

Omega coefficients for the bifactor model using WLSMV estimation ranged from .96 for the School Climate general factor to .72 for the Physical Safety subfactor, with all other subfactor Omega coefficients greater than .80.

Thus, the MLR pattern of results was confirmed using WLSMV estimation. The theoretical GSCS yielded better fit than the original GSCS, and the bifactor model fit the data best. Across estimation methods, the GSCS items were generally good indicators of the general school climate factor, and slightly more total variance could be attributed to the general factor than to the specific factors. Lastly, the GSCS factors showed acceptable reliability across estimation methods. The only appreciable differences across estimation methods were that: (1) RMSEA indicated slightly worse fit using WLSMV estimation, and CFI/TLI slightly better fit, across each model; (2) factor loadings were slightly higher using WLSMV estimation, and (3) Omega coefficients were slightly higher using WLSMV estimation.

Step 2: Estimation of variance and reliability. The ICCs were calculated to examine the appropriateness of a multilevel analysis. The ICC(1) indicates the proportion of the total variance in an item that can be explained by the school level. Using WLSMV estimation, ICC(1) estimates ranged from .02 (Items 1, 2, 4, and 7) to .21 (Item 36). Twenty out of 30 items had ICC(1)s of .05 or greater, supporting the use of MLCFA. The ICC(2) coefficient was calculated to examine the reliability of the group (within school) averages on items. The ICC(2) indicates the degree to which students within schools perceive school climate similarly. Using WLSMV estimation, ICC(2) estimates ranged from .92 (Item 4) to .99 (Items 28, 29, 30, 36, and 38), indicating a high level of agreement among students within the same school and suggesting the GSCS is a reliable estimate of school-level constructs.

Thus, the MLR pattern of results was confirmed using WLSMV estimation. The ICC(1) and ICC(2) coefficient results followed a similar pattern, and supported the use of MLCFA. The only appreciable differences across estimation methods were that The ICC(1) and ICC(2) coefficient results were slightly greater using WLSMV estimation.

Step 3: Multilevel factor analysis. This step involved testing the measurement structure within and between groups. Note that a bifactor model with one general factor and seven specific factors and with item indicators for factors measured both within and between groups is a complex model, especially with items designated as ordered-categorical indicators. Nevertheless, using WLSMV estimation, this model represented a good fit for the data (e.g., CFI = .97, TLI = .97, and RMSEA = .03; see Table A1). For the general School Climate factor, all item loadings, except for the school attachment items, were greater for the between-schools portion of the model than for the within-school portion. Loadings ranged from −.04 (item 2) to .95 (item 89) for the between-schools model (M = .80) and ranged from .24 (item 32) to .69 (items 12, 13, and

14) for the within-school model (*M* = .54). For School Attachment factor, loadings were lower for the within-school model, while for all other specific factors, loadings were generally greater for the within-school model.

Thus, the MLR pattern of results was again confirmed using WLSMV estimation. Across estimation methods, fit statistics indicated good fit and confirmed that the model fit the data at both the within and between levels, and the pattern of factor loadings was consistent across estimation methods. The only appreciable differences were similar to those described above (e.g., CFI and TLI were slightly higher using WLSMV estimation, as were factor loadings). Because MLR and WLSMV estimation had produced similar results thus far, the remaining analyses were conducted using MLR estimation to reduce computation time and to provide specific fit indices for each level (i.e., SRMR Within and Between).

Table A1.
WLSMV Estimation Fit Statistics

	χ^2	df	CFI	TLI	RMSEA
GSCS Models					
Original GSCS	677978.4*	566	.932	.925	.068
Theoretical GSCS	354767.5*	384	.961	.956	.060
Theoretical GSCS Alternative Models (Single Level)					
One Factor	1829632.2*	405	.798	.783	.132
Multifactor	354767.5*	384	.961	.956	.060
Second Order	359652.8*	398	.960	.957	.059
Bifactor	308694.6*	375	.966	.961	.056
MLCFA					
Multilevel Bifactor	137865.3*	750	.976	.973	.027

CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Square Error of Approximation.
*p<.001.

Appendix B

Supplement to Data Analysis Steps 1-6

Step 1: Conventional factor analysis. First, a conventional confirmatory factor analysis (CFA) was conducted to fit the proposed factor structures to the total sample covariance matrix. The original GSCS factor structure (see Table 2.1) and the theoretical GSCS factor structure (see Table 2.2) were specified. The CFA aims to explain the covariance among the observed variables (p items) in terms of the latent factors (q factors):

$$Y_{pi} = \nu_p + \lambda_{pq}\eta_{qi} + e_{pi} \tag{2}$$

where, for individual i, Y_{pi} is the score for the pth observed dependent variable (item); η_{qi} is the score for the qth latent factor; λ_{pq} is the regression (factor loading) for the pth item on the qth factor, which indicates the magnitude of change in the item for each unit change in the factor; ν_p is the regression intercept; and e_{pi} is the observed variable residual score. Results of each model were analyzed to determine which factor structure best fit the data and whether revisions to the model were indicated.

Next, alternative factor structures were specified, including a unidimensional model (in which all items load onto a single factor), a second-order model (in which the latent school climate sub-factors serve as indicators of an overall school climate factor), and a bifactor model (in which item variance is partitioned between a general school climate factor and specific factors, and the specific factors represent the shared variance of items not accounted for by the general school climate factor). Model fit statistics of the alternative factor structures were compared to those of the multi-factor model to determine the best-fitting structure. While conventional (single-level) CFA is inappropriate when analyzing data from students nested within schools (due to the correlated observations of clustered data), single-level factor analysis of clustered data often produces inflated model fit. Thus, the aim of these analyses was to identify obvious model misspecifications (Muthén, 1994).

Step 2: Estimation of variance and reliability. The intra-class correlation (ICC) coefficients were calculated to examine the appropriateness of a multilevel analysis (Hox 2010; Muthén, 1994, Muthén and Satorra, 1995). The ICC(1) indicates the proportion of the total variance in an item that can be explained by the school level as an estimate of:

$$\frac{{\sigma_B}^2}{{\sigma_B}^2+{\sigma_W}^2} \tag{3}$$

where σ_W is the pooled-within variance, a consistent and unbiased estimator of item variance within schools, and σ_B is the variance between schools, weighted by the group size. Small ICC(1)s would suggest the majority of the variance in student responses on the GSCS is explained by individual differences within schools. Larger ICC(1)s would indicate that the variability is explained by differences between schools and provide additional support for the necessity of a multilevel factor analysis. For reference, ICC(1)s for school performance measures (e.g. reading or math) typically range from .1 to .25 in American schools (Hedges & Hedberg, 2007). However, ICC(1)s for nonperformance school measures, such as school climate, are often more modest (Morin et al., 2014). Any ICC value above 0 suggests the data dependency, and thus supports multilevel modeling (Bliese, 2000). Monte Carlo simulations have shown that ICC(1)s > .05 indicate the need to analyze the data using multilevel CFA (Julian, 2001).

Additionally, the ICC2 coefficient was calculated to examine the reliability of the group (within school) averages on items (Morin et al., 2014). The ICC(2) indicates the degree to which students within schools perceive school climate similarly and is computed as follows:

$$\frac{{\sigma_B}^2}{{\sigma_B}^2+({\sigma_W}^2/n_j)} \tag{3}$$

where n_j is the average size of the groups. Since school climate is inherently a school-level variable, in an optimal scenario, there would be a high level of agreement among students within the same school. A complete lack of agreement may indicate that the instrument is not a reliable estimate of the school level construct (Morin et al., 2014).

Step 3: Multilevel factor analysis. Multilevel confirmatory factor analysis (MCFA) allows simultaneous estimation of the within-level (students within schools) and between-level (between schools) measurement models. The MCFA breaks down the total sample covariance matrix into a pooled-within group covariance matrix and between-group covariance matrix, using these two matrices to estimate the factor structure at each level and allowing level-1 indicator intercepts to be random:

$$Y_{ij}^{P} = \nu_j^{P} + \lambda_{pq}^{(w)} \eta_{ij}^{Q(w)} + e_{ij}^{P(w)} \quad (4)$$

$$\nu_j^{P} = \nu^{P} + \lambda_{pq}^{(b)} \eta_{j}^{Q(b)} + \xi_{j}^{P(b)} \quad (5)$$

where the first equation estimates the within group part of the model and the second equation estimates the between group part of the model. These equations are linked by the school-specific regression intercept ν_{jP}, which also serves as the dependent variable at the between level. In other words, the regression intercept of each indicator is a random variable at the between level – it can vary across schools. The between level latent factor $\eta_{iQ(b)}$ accounts for the variance of the school-specific intercepts, and $\xi_{jP(b)}$ is the remaining residual at the between level after accounting for the school level latent factors.

Step 3a: Multilevel confirmatory factor analysis. An MCFA was conducted with the final factor identified in *Step 1* specified at both the within and between levels. Results were analyzed to determine whether the factor structure fit the data at each level and whether revisions to the factor structure were warranted.

Step 3b: Cross level measurement invariance. To ascertain whether school climate is being measured the same way at the individual-level and the school-level, cross-level invariance procedures were employed. A series of MCFA models were specified and examined to determine if the factor structure identified in *Step 1* held at each level.

First, configural models were examined to determine if the number of factors and pattern of factor loadings was the same across levels of analysis (i.e. *configural invariance*). Because the factor structure of single-level models is largely influenced by the within-school data (i.e., student level data) due to larger sample sizes at this level, the between-schools level was investigated to explore whether alternative structures better explained the school-level data. With the within-school model fully saturated, alternative models were specified at the between-schools level (i.e., unidimensional, multi-factor, second order, and bifactor models) and evaluated in terms of model fit. Next, the pattern and rank order correlations of factor loadings were examined across levels. If the factor structure and pattern of factor loadings is similar across levels, cross-level configural invariance can be assumed across the individual and school level.

If support for configural invariance is found, the next step is to specify a metric model to test the hypothesis of equality of factor loadings across levels (i.e. *metric invariance*). If constraining the factor loadings to be the same across levels does not worsen the model fit, there is reasonable support that the constructs are similarly measured at the individual-level and the school-level and latent factor variances can be directly compared. Thus, it can be determined how much variability in the latent factors is due to the individual or school level (Mehta and Neale, 2005). If constraining the factors loadings across levels does significantly worsen the model fit, partial metric invariance can be explored by examining the modification indices to determine the source of non-invariance. Suggested factor loadings can be freed one at a time until model fit is appropriate. In contrast to testing invariance across groups, support for configural and metric invariance is considered sufficient to determine cross-level invariance of factor structures (i.e. scalar invariance is not tested).

Step 3c: Cluster bias. Cluster bias was examined by constraining the final metric model so that the item error variances at the between level are zero (Jak et al., 2013). If this additional constraint does not significantly worsen model fit, it can be assumed that all the between-level variability in the items is explained by the school-level climate latent factors and that the indicator intercepts are equal across schools. If the constraint substantially worsens model fit, modification indices can be examined to determine the presence of cluster bias in specific items (Jak et al., 2013). Suggested items can be freely estimated, one-by-one, until model fit is appropriate. The presence of cluster bias in an item implies that item does not measure school climate equally across schools. That is, a student's response to that item not only depends on the student's perception of the corresponding school climate factor, but also on unaccounted for school characteristics, such as school size or location. Therefore, psychometrically sound comparisons of group means across schools are not valid. If cluster bias is detected, school-level factors can be explored to determine if they account for the variability. Reliability estimates of the final MCFA model were calculated using procedures outlined in Rodriguez et al. (2016).

Step 4: Individual-level group measurement invariance. Testing the measurement invariance of multilevel models in relation to within-level groups (i.e. racial/ethnic groups) is not feasible using traditional multiple-group CFA procedures because the group indicators (e.g. male and female students within schools) are crossed at the between level (across schools). To address this, the present study employed recommendations from Kim et al. (2015) and Jak (2013) to use

ML MIMIC modeling procedures. Thus, the individual-level race/ethnicity variables (grouping variables) were added to the model, which allowed testing group differences via a regression-type analysis.

To test for measurement invariance using ML MIMC modeling procedures outlined in Kim et al. (2015), sets of nested models are compared in terms of model fit. A constrained reference model is constructed in which the grouping variable is added with direct effects specified on the corresponding within-level latent school climate factors (to test for group differences in terms of a latent factor). Then, a relaxed model is created with *two additional direct effects* specified: one from the grouping variable on the within-level indicator of interest (to test for intercept invariance), and one from an interaction variable (between the grouping variable and the corresponding latent factor) on the within-level indicator of interest (to test for factor loading invariance). The relaxed model is compared to the constrained model (with the two additional direct effects constrained to be zero) in terms of fit statistics and direct effects.

The procedure outlined in Kim et al. (2015) to test for factor loading invariance using ML MIMIC modeling produced convergence issues in the present study. Other studies have shown that this procedure leads to convergence problems and biased parameter estimates when the observed indicator is not normal (Bagheri et al., 2018; Cham et al.). Thus, traditional ML MIMIC modeling procedures were employed following procedures outlined in Jak (2013). To investigate the assumption of intercept invariance, a constrained model is constructed in which the race/ethnicity grouping variable is added with direct effects specified on the corresponding within-level latent school climate factors and with direct effects on the within-level indicators constrained to zero. The fit of this model is evaluated, and modification indices are examined to determine the presence of non-invariant items. If freeing suggested direct effects does not improve model fit, within-level intercept invariance can be assumed. If freeing a suggested direct effect does improve model fit, and if the direct effect is significant, it suggests the item's intercept is not invariant across racial/ethnic groups.

Traditional ML MIMIC (i.e., comparing a constrained model in which the grouping variable is added with a direct effect specified on the latent variable to a relaxed model in which an additional direct effect from the grouping variable to the factor indictor of interest is specified) assume nonuniform invariance of scales (i.e., factor loading invariance). To investigate the plausibility of this assumption, the TYPE=COMPLEX function was utilized in Mplus and multiple-group invariance procedures were employed. Results provided support for configural and metric invariance of the factor structure across racial/ethnic groups (i.e., Asian/Pacific-Islander, Black/African American, European American/White, Hispanic/Latino, and Other). The configural model fit the data well (e.g., RMSEA = .04, CFI = .94, TLI = .93, and SRMR = .04), and constraining the factor loadings did not significantly worsen model fit (e.g., ΔRMSEA = −.002, ΔCFI = −.001, ΔTLI = .006, and ΔSRMR = .007) according to criteria specified in Cheung and Rensvold (2002) and Chen (2007).

Step 5: School-level measurement invariance. To test for measurement invariance of the multilevel model in relation to school-level groups (e.g. school racial/ethnic and SES composition), multilevel MIMIC modeling procedures were employed. MIMIC modeling procedures allow for invariance testing in reference to continuous variables and allow for between-level measurement invariance to be estimated in a step-wise fashion, accounting for the effect of the within-level grouping factors (Jak, 2013). ML MIMIC modeling invariance testing procedures were the same as described for the individual-level factors, except paths were specified at the between level. Multiple-group invariance procedures could not be employed due to the continuous nature of the grouping variable. Thus, only intercept invariance (i.e., uniform invariance) was tested. Researchers have stated that uniform noninvariance is more important than non-uniform noninvariance (i.e., factor loading noninvariance), because noninvariance of factor loadings often evens out at the scale level (Huang et al., 2011).

Step 6: Analyses of relationships. The final model, with each student- and school-level grouping variable included with appropriately specified paths (as determined in *Steps 5* and *6*), was examined to determine the relationships between the grouping variables and the overall school climate factor and its sub-factors.

Before structural relationships were interpreted, latent factor reliability estimates of the final measurement model from *Step 3* were calculated using procedures outlined in Rodriguez et al. (2016) to determine whether estimated relationships between latent factors and grouping variables could be considered trustworthy. Coefficient omega (ω; McDonald, 1999) is a factor analytic model-based internal reliability estimate. Coefficient omega hierarchical (omegaH or ω_H) estimates the proportion of variance in total scores that can be attributed to the general factor. When omegaH is high (>.80), total scores can be considered essentially unidimensional (Reise et al., 2013). Coefficient omega hierarchical subscale (omegaHS or ω_{HS}) reflects the reliability of a subscale score after controlling for the variance due to the general factor (Reise et al., 2013). Relative omega for a general factor represents that proportion of reliable variance attributable to the general factor. For specific factors, it represents the proportion of reliable variance in the subscale after partitioning out the variance for the general factor. Of particular relevance for the analysis of relationships with other variables, the *H* index provides a measure of construct reliability that estimates the quality of a latent factor's indicators, and, thus, its replicability across studies (Hancock, 2001). A low *H* value suggests "the latent variable is not well defined by the indicators and, thus, is expected to change across studies;" whereas, a high *H* value (>.70) suggests a well-defined latent variable that "will have more stability across studies" (Rodriguez et al., 2016, p. 143).