IES Institute of Education Sciences

Strengthening the Pennsylvania School Climate Survey to Inform School Decisionmaking

Regional Educational Laboratory Mid-Atlantic

At Mathematica

August 2024
REL 2024-006
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation and Regional Assistance at IES

The Institute of Education Sciences (IES) is the independent, nonpartisan statistics, research, and evaluation arm of the U.S. Department of Education. The IES mission is to provide scientific evidence on which to ground education practice and policy and to share this information in formats that are useful and accessible to educators, parents, policymakers, researchers, and the public.

We strive to make our products available in a variety of formats and in language that is appropriate for a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other IES product or report, we would like to hear from you. Please direct your comments to ncee.feedback@ed.gov.

This report was prepared for IES under Contract 91990022C0012 by Mathematica. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

August 2024

Burnett A., Milless, K. L., Bennett, M., Kozakowski, W., Alves, S., & Ross, C. (2024). *Strengthening the Pennsylvania School Climate Survey to inform school decisionmaking* (REL 2024-006). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. https://ies.ed.gov/ncee/rel/Products/Region/midatlantic/Publication/108114.

This report is available on the Institute of Education Sciences website at https://ies.ed.gov/ncee/rel/.

Strengthening the Pennsylvania School Climate Survey to Inform School Decisionmaking

Alyson Burnett, Katlyn Lee Milless, Michelle Bennett, Whitney Kozakowski, Sonia Alves, and Christine Ross

August 2024

This study analyzed Pennsylvania School Climate Survey data from students and staff in the 2021/22 school year to assess the validity and reliability of the elementary school student version of the survey; approaches to scoring the survey in individual schools at all grade levels; and perceptions of school climate across student, staff, and school groups. The survey encourages data-informed efforts in participating Pennsylvania schools to foster supportive learning environments that promote social and emotional wellness for students and staff. The study validated the elementary school student survey but found that one domain—safe and respectful school climate—did not meet the reliability threshold and thus suggests that revisions are needed. At all grade levels noninstructional staff had the most positive perceptions of school climate, followed by classroom teachers then students. The study found that different approaches to combining the school climate scores of students, teachers, and noninstructional staff within schools yielded slightly different distributions of school climate summary index scores. It also found that different performance category thresholds resulted in similar distributions of schools across categories. Scores calculated using simple averages were strongly and positively correlated with scores calculated using a more complex approach (Rasch models), suggesting that both approaches deliver similar information. School climate scores varied across student groups (defined by race/ethnicity, gender, and grade level) within schools and across school groups. Larger schools and schools with higher percentages of Black students tended to have lower school climate scores than other schools. The findings can inform the Pennsylvania Department of Education's decisionmaking on revisions to the elementary school student survey, approaches to scoring and reporting survey results, and efforts to increase participation in future survey administrations.

Why this study?

The Pennsylvania Department of Education's (PDE) Office for Safe Schools aims to help schools across the state foster positive school climates that promote social and emotional wellness for students and staff. School climate refers to the quality and character of school life, encompassing social, emotional, and physical aspects of the school environment. It describes the interpersonal relationships among members of a school community; its teaching and learning practices; and its values, norms, and structures (National School Climate Council, 2007). A positive school climate suggests healthy and rewarding relationships among students and school staff and a sense of safety, respect, and engagement in teaching and learning. School climate has the potential to influence students' social and academic development (Cohen et al., 2009; Voight & Hanson, 2017). Specifically, school climate can affect students' social and emotional wellness, with a positive school climate offering better opportunities for students to succeed and thrive (Aldridge & McChesney, 2018). Students in schools with more positive climates tend to have higher average academic performance and positive mental health outcomes, perhaps because school climate captures information about how supportive the learning environment is, which can affect student learning (Daily et al., 2019; Kwong & Davis, 2015; Voight & Hanson, 2017). School climate

For additional information, including technical methods, supplemental findings, and additional findings, access the report appendices at https://ies.ed.gov/ncee/rel/Products/Region/midatlantic/Publication/108114.

The authors acknowledge everyone who contributed to this report, including Charles Tilley, Jensen Hu, and Evan Morier for data analytics; Brian Gill, Philip Gleason, and Matthew Steinberg for expert review; Yange Xue, Joshua Stewart, and Tim Kautz for consultation on psychometric analyses; Jennifer Brown for editing; and Sheryl Friedlander for report production.

can also shed light on the social and emotional wellness of staff, with positive climates associated with greater teacher satisfaction, lower stress and burnout, and higher teacher retention (Cohen et al., 2009; Collie et al., 2012; Grayson & Alvarez, 2008).

The PDE Office for Safe Schools administers the Pennsylvania School Climate Survey to schools that choose to participate, analyzes the scores, and provides each school with a score report. In the 2021/22 school year, 11 percent of Pennsylvania public schools participated in the survey. Five different surveys measure school climate for elementary school students (typically grades 3–5), middle school students (typically grades 6–8), high school students (typically grades 9–12), classroom teachers, and noninstructional staff (school staff who are not classroom teachers, including administrators, cafeteria staff, custodians, and others who have contact with students). Having a variety of respondents allows for exploring differences and similarities in experiences. The scores for all five respondent types are reported on three domains (see box 1 for definitions of key terms used in the report), aligned with three dimensions of school climate that most interest PDE: social-emotional learning, student support and academic engagement, and safe and respectful school climate.

School climate surveys have great potential to improve understanding of students' and staff's perceptions of their school environment—including the extent of support for social-emotional learning, the level of emotional and academic support, and the sense of safety and respectfulness in the school (Aldridge & McChesney, 2018; Kwong & Davis, 2015; Voight & Hanson, 2017). Once the validity and reliability of the Pennsylvania School Climate Survey (and each of its domains) have been established, PDE can use the survey to describe school climate within and across schools, over time, and across respondent characteristics. Schools and districts can use the survey results to evaluate and track their efforts to improve school climate over time.

In a previous partnership between the Regional Educational Laboratory Mid-Atlantic and PDE, a study team investigated the validity and reliability of the Pennsylvania School Climate Survey and developed an approach for calculating a school-level summary score across survey items and respondents using Rasch models (Amos & Xue, 2021). That effort established the validity and reliability of the middle and high school student surveys, the teacher survey, and the noninstructional staff survey and identified changes that would improve the validity and reliability of the elementary school student survey. The current study reassesses the validity and reliability of the elementary school student survey after the recommended changes were implemented.

The study also provides updated guidance on how to combine the responses of students, teachers, and noninstructional staff into a single school-level score (called the school climate summary index score) for use across school levels. Because calculating scores using Rasch models requires specialized computer programs and expertise, PDE is interested in a mean-based approach that does not require additional support. For this reason the study assesses the extent to which scores calculated using Rasch models and mean-based scores provide the same information. The study also explores ways of defining performance categories and assesses differences in scores by respondent characteristics.

PDE plans to use the findings to expand the information provided in its school climate score reports to make them more useful and actionable for schools. For example, under PDE's current approach to defining categories, nearly all schools fall into the middle performance category, which makes it difficult for schools to recognize a need for improvement. Performance categories that allow schools to understand whether their school climate needs to improve overall or in specific areas could make the score reports more useful. By improving information in the score reports, PDE hopes to encourage widespread use of the survey.

Given that school climate surveys are widely used throughout the United States, the findings from this study may also be useful for other states and school districts that are validating surveys and refining scoring procedures. States and districts may wish to follow the steps this study took to assess their survey's validity and

reliability. They may also be pondering similar questions about scoring their surveys, including whether to use Rasch models or mean-based scores, how to weight different types of respondents, and how to create performance categories. The discussion of the advantages and disadvantages of various scoring approaches could help inform decisions that make school climate scores more useful in states and districts beyond Pennsylvania.

Box 1. Key terms

Survey domain. A group of items in a survey that intend to measure a single topic within the overall topic of the survey (sometimes a survey domain is called a scale). The Pennsylvania School Climate Survey has three domains:

- *Social-emotional learning.* Measures respondents' perceptions of students' social and emotional regulation and conflict resolution skills. For example, this domain includes the item "Students in my school stop and think before doing anything when they get angry" in the student surveys.
- *Student support and academic engagement.* Measures respondents' perceptions of how much students are listened to, cared about, and helped by teachers and other adults in the school. For example, this domain includes the item "My teachers really care about me" in the student surveys.
- *Safe and respectful school climate.* Measures respondents' perceptions of students' and teachers' physical and emotional safety. For example, this domain includes the item "Students in my school treat each other with respect" in the student surveys.

School climate index score. A score that combines all school climate survey domain scores. It can be calculated for each individual, for each respondent type (e.g., the average school climate index score for teachers within a school), or across all respondents in a school (see "summary score" below).

Summary score. A school-level score that combines the average scores of all respondent types (students, teachers, and noninstructional staff). Each school has a summary score for each survey domain and for the index.

Performance category. A classification of the school's climate based on a school climate index or domain score. Performance categories translate scores into a more understandable rating. These categories can be labeled numerically (e.g., 1–4) or qualitatively (e.g., "Unfavorable," "Fair," "Favorable," and "Most favorable").

Validity. The extent to which a measure, such as a survey, captures what it is designed to measure. Different forms of validity provide different types of evidence that the survey captures what it intends to measure. This study assessed **construct validity** (the extent to which the items in a survey or survey domain measure the specific construct, or topic, that the domain claims to measure) through **convergent validity** (the extent to which survey items are closely related to each other and to the underlying construct) and **discriminant validity** (the extent to which separate surveys or survey domains capture different underlying constructs).

Reliability. The extent to which a measure, such as a survey domain, consistently produces the same results of an underlying construct, such as the quality of student-teacher relationships. One measure of reliability is Cronbach's alpha, which measures how related a set of survey items are to each other.

Mean-based scores. Scores calculated using a simple mean across all survey items in a domain (for an individual respondent's domain score) and a simple mean across domain scores (for an individual respondent's school climate index score). Individual domain and index scores can then be averaged together to create school-level scores.

Rasch models. Psychometric models that estimate a respondent's likelihood of selecting a response option for each survey item as a function of the overall construct (e.g., a respondent's perception of school climate). Rasch models can account for different response patterns, such as different distances between "agree" and "strongly agree" across items (Dogan, 2018).

Effect size. The size of the difference between two means, measured in this study using Cohen's *d*. It is calculated as the difference in two means divided by the pooled standard deviation.

Research questions

The study addressed eight research questions across three topic areas using the data from the Pennsylvania School Climate Survey administered during the 2021/22 school year:

1. Validity and reliability of the elementary school student survey.
   a. To what extent is the elementary school student survey valid? How does the validity of the elementary school student survey compare with the validity of the middle and high school student surveys?
   b. To what extent is the elementary school student survey reliable? How does the reliability of the elementary school student survey compare with the reliability of the middle and high school student surveys?

2. Approaches to scoring the school climate survey in individual schools.
   a. To what extent did perceptions of school climate vary by respondent type (student, teacher, or noninstructional staff)?
   b. How does the approach to weighting the responses of students and staff in a school affect the distribution of school climate summary index scores among schools?
   c. How does the distribution of schools across performance categories (such as on a scale from 1, least favorable, to 4, most favorable) compare when thresholds are mapped to the survey scales, are based on Rasch models, and are based on percentiles?
   d. What is the relationship between scores calculated using Rasch models and scores calculated using averages on the four-point survey scale (mean-based scores)?

3. Perceptions of school climate across student and school groups.
   a. To what extent did perceptions of school climate vary by students' race/ethnicity, gender, and grade?
   b. To what extent did perceptions of school climate vary by school characteristics—namely, school size, student racial/ethnic composition, percentage of students eligible for the National School Lunch Program, average student achievement, and school urbanicity (urban, rural, suburban, or town)?

Box 2 summarizes the data sources, sample, methods, and limitations, and appendix A provides technical details.

Box 2. Data sources, study sample, methods, and limitations

Data sources. This study used three data sources from the 2021/22 school year:

- Pennsylvania School Climate Survey responses from students, teachers, and noninstructional staff in schools across Pennsylvania that opted to participate in the survey.
- Publicly available administrative data from the Pennsylvania Department of Education (PDE) on statewide counts of public school teachers and noninstructional staff and schools' proficiency rates in reading/language arts and math.
- Administrative data on school characteristics from the Common Core of Data, including school size, demographics of students attending the school, and school urbanicity.

Study sample. This study included 301 elementary, middle, and high schools that opted to administer the Pennsylvania School Climate Survey during the 2021/22 school year.[1] They represent approximately 11 percent of all Pennsylvania public schools. Schools that had at least 10 student respondents, 5 teacher respondents, or 5 noninstructional staff respondents were included (see appendix A for the number of schools dropped for each restriction).

Survey response rates. In each school an average of 68 percent of students and 56 percent of teachers responded to the survey. Response rates for noninstructional staff could not be calculated because school-level counts of noninstructional staff were not available (see tables A8–A10 in appendix A for more detail and response rates by school level).

The study team tested whether students who completed the surveys were representative of their schools' populations based on intersectional racial/ethnic and gender groups.[2] Multiracial students tended to be underrepresented, whereas White students and Hispanic students tended to be overrepresented (see table A12 in appendix A for details and rates of imbalance for all student groups). Nonresponse weights were applied to student scores to account for differences between the demographic composition of the survey respondents and that of their school's population.[3] The study team could not assess nonresponse bias for teachers and noninstructional staff due to lack of demographic data for these groups.

Generalizability. The study team assessed whether schools in the sample were representative of the remaining public schools in the state. Schools in the sample tended to have fewer Asian or Pacific Islander students and Black students but more Hispanic students and White students than other public schools in the state. In addition, the sample included a smaller percentage of urban schools and a larger percentage of schools in towns than the remaining public schools in Pennsylvania (see tables A4–A7 in appendix A).

Methods. The study used the following methodological approaches for each research question:

Research question 1a. To assess construct validity, or how well the survey and its domains measure the topics they intend to measure, the study team used confirmatory factor analysis. This involved fitting a model and assessing overall fit statistics, which summarize whether a structural equation model fits the data well; standardized factor loading estimates, which indicate the strength of the associations between each survey item and the underlying construct; and correlations between factors, which provide evidence of discriminant validity, or whether the survey domains capture different underlying constructs. See appendix A for the thresholds of these estimates used to assess validity. Only survey items and domains that met the specified construct validity criteria are included in the subsequent analyses.

Research question 1b. To assess reliability, or the extent to which the survey consistently produces the same results, the study team calculated Cronbach's alpha, a measure of internal consistency, which shows whether the items in a domain produce consistent results. The team evaluated whether the Cronbach's alpha was .70 or higher, a commonly used criterion for reliability (Bland & Altman, 1997). Only domains that met the threshold for reliability are included in subsequent analyses.

Research question 2a. The study team conducted the same validity and reliability analyses used for the student surveys in research questions 1a and 1b to ensure validity and reliability of the teacher and noninstructional staff surveys from the same school year (see appendix C). Next, the study team calculated scores for each respondent. Each item included a four-point response scale ranging from 1 (strongly disagree) to 4 (strongly agree). To create an individual respondent's score on a given school climate domain (social-emotional learning, safe and respectful school climate, and student support and academic engagement), the study team calculated a simple mean across survey items belonging to the domain. These are referred to as "mean-based scores" throughout the report. The three school climate domain scores were averaged together to create a school climate index score for each respondent. Individual respondents' scores within each school and respondent type were then averaged together (after nonresponse adjustments were applied to student scores) to create a school-level score for each of three school climate domains and for each respondent type (students, teachers, and noninstructional staff).

Using paired sample *t*-tests, the study team summarized the within-school differences among different respondent types—testing whether students, teachers, and noninstructional staff in the same school tended to give similar or different assessments of the school's climate. In addition, the study calculated effect sizes using Cohen's *d* to evaluate the magnitude of differences on a metric that would be comparable to those in other studies. Based on prior studies of school climate (Kautz et al., 2021), the current study described effect size differences as follows: an absolute value of 0.0–0.09 is small, 0.10–0.19 is moderate, 0.20–0.29 is substantive, and 0.30 and above is large.

The study team also calculated correlations of school climate scores between students, teachers, and noninstructional staff. To describe the strength of these correlations, the study team used guidance from Cohen (1988): correlations below .30 are low, correlations of .30–.49 are moderate, and correlations of .50 and above are strong.

Research question 2b. Using mean-based scores, the study team calculated school-level summary scores that combined scores across all three respondent types. The team calculated summary scores for each school climate domain and for the school climate index. The team assessed two approaches to weighting respondent types to create the school climate summary scores: group weighted, which equally weighted students and staff (teachers and noninstructional staff, who were weighted equally to one another), and state weighted, which weighted each respondent type according to their population share in Pennsylvania. The second approach gives students more weight (.87) than the first approach (.50) but ensures that each person is implicitly equally weighted. The team compared the distributions of scores using both weighting approaches.

Research question 2c. The study team compared the distributions of schools across performance categories under different approaches to defining those categories. The study team defined thresholds for each performance category that represent the upper and lower boundaries of school-level scores within each category using four approaches (see table A14 in appendix A):

- The current PDE approach, which maps performance categories to the survey scale consistently across domains and respondents. Scores less than or equal to 2.5 are categorized as 1 (least favorable), scores greater than 2.5 and less than 3.4 are categorized as 2, and scores greater than or equal to 3.4 are categorized as 3 (most favorable).
- A modified PDE approach, developed by the study team, that maps performance categories to the survey scale consistently across domains and respondents. It is similar to the current PDE approach but with four performance categories rather than three: scores less than or equal to 2.5 are categorized as 1 (least favorable), scores greater than 2.5 and less than or equal to 3.0 are categorized as 2, scores greater than 3.0 and less than or equal to 3.5 are categorized as 3, and scores greater than 3.5 are categorized as 4 (most favorable).
- A Rasch-identified approach that is based on Rasch models (which estimate a respondent's likelihood of selecting a response option for each survey item as a function of the respondent's perception of school climate) for each school climate domain within each respondent type. Rasch models can identify step values, or points at which respondents shift from being more likely to select an answer such as "strongly agree" to being more likely to select another answer such as "agree." The team used the step values from the Rasch models to define thresholds for different ranges of scores then translated the thresholds to the survey scale units.
- A reference-based approach that allows individual schools to compare their school climate scores to the distribution of all schools in the state that participate in the survey. The study team created performance categories mapped to quartiles where schools in the bottom quartile are categorized as 1 (least favorable) and schools in the top quartile are categorized as 4 (most favorable) for each domain and respondent type.

The study team then calculated the distribution of schools across performance categories for domain scores and school climate index scores. The team did this for each respondent type separately and across respondents using the group weighted and state weighted approaches. The team compared the distributions of schools across performance categories under the first three approaches (the fourth approach is not compared because 25 percent of schools are always in each category).

Research question 2d. The study team assessed the relationship between mean-based scores on the 1-4 survey scale and scores based on Rasch models. The team calculated correlations between mean-based scores and Rasch scores for each domain and for the school climate index. These correlations were calculated for each respondent type (students, teachers, and noninstructional staff) separately and for the school-level summary scores that combine results across respondent types. The study team considered correlations above .90 to be very strong because they show that more than 90 percent of the variation of one type of score is explained by the other type of score.

Research question 3a. The study team conducted a series of paired sample *t*-tests to assess differences in perceptions of school climate between students of different races/ethnicities, genders, and grades within schools. As with research question 2a, these *t*-tests also adjusted the *p*-values to account for multiple comparisons. The study team also calculated effect sizes to characterize the magnitude of differences.

Research question 3b. To identify whether certain school characteristics are related to school climate scores, the study team calculated correlations between school climate scores and school characteristics, such as school size, racial/ethnic composition and socioeconomic status of students, proficiency rates for reading/language arts and math, and school urbanicity. The study team characterized the strength of the correlations using the same guidance as in research question 2a.

Limitations. Schools in the analysis sample participated in the survey on a voluntary basis and are not representative of all schools in the state, so the findings may not apply to all schools in Pennsylvania. Furthermore, due to nonresponse, the students within a school who respond to the survey are not representative of the school's overall student population.[4] Although nonresponse weights help address this limitation, the study team could not account for all potential differences between respondents and nonrespondents. Additionally, nonresponse weights were not available for staff or noninstructional staff, so it is not known whether the analyses are representative of these groups. See appendix A for more discussion of the study's limitations.

Notes

1. The survey was publicized by PDE, and individual schools had to sign up to administer it.

2. This analysis included 253 schools with at least 10 student respondents; demographic information for teachers and noninstructional staff were not available, and therefore nonresponse bias analyses could not be conducted for teachers or noninstructional staff.

3. Nonresponse weights for teachers and noninstructional staff were not calculated because data on their population demographics at the school level were unavailable.

4. Nonresponse may be due to individual students who opted not to take the survey or who were absent from school on the day it was administered or due to teachers who opted not to administer the survey to their classes. Because survey responses did not include classroom-level identifiers, the study team could not confirm whether nonresponse was driven more by individual- or classroom-level nonresponse.

Findings

This section presents findings to address the study's research questions. Detailed supplemental findings are presented in appendix B.

Validity and reliability of the elementary school student survey

The study assessed the validity and reliability of all three domains for the elementary school student survey and compared the results with those for the middle and high school student surveys.

The elementary school student survey demonstrated validity, and domains were acceptably differentiated from one another. All domains in the elementary school student survey demonstrated construct validity and discriminant validity (see tables B1–B14 in appendix B for fit statistics, factor loadings, and correlations). To meet the construct validity criteria, the study team dropped two items from the safe and respectful school climate domain with factor loadings below .40 ("Students at my school are bullied" and "Students at my school are teased, picked on, made fun of, or called names"). After these items were dropped, the remaining set of items met the thresholds for acceptable construct validity, indicating that they measured the domains as intended (see table B9).

The correlations between domains all met the threshold for establishing discriminant validity (.85 or lower), indicating that they measure different underlying constructs. However, the correlation between the safe and respectful school climate domain and student support and academic engagement domain was very close to the threshold for high correlation (.84; see table B14 in appendix B).[1]

1. The construct validity and discriminant validity of the elementary school student survey were comparable to those of the middle and high school student surveys after the two items mentioned above were removed.

Two domains in the elementary school student survey (social-emotional learning, and student support and academic engagement) demonstrated acceptable reliability, and one domain (safe and respectful school climate) failed to meet the threshold for acceptable reliability. The social-emotional learning domain and the student support and academic engagement domain demonstrated acceptable reliability on the elementary school student survey, meaning they had a Cronbach's alpha of .70 or higher (figure 1; Bland & Altman, 1997).

The safe and respectful school climate domain, with a Cronbach's alpha of .63, did not meet the threshold for acceptable reliability. This may be due to the small number of items retained as valid (four, after two were removed). Reliability may also be low because three of the valid items in the domain relate to a sense of safety, while only one relates to the perception of respect among students. In contrast, all domains for the middle and high school student surveys and the teacher and noninstructional staff surveys met the reliability criteria. See table B15 in appendix B for the Cronbach's alpha for each combination of survey domain and respondent type.

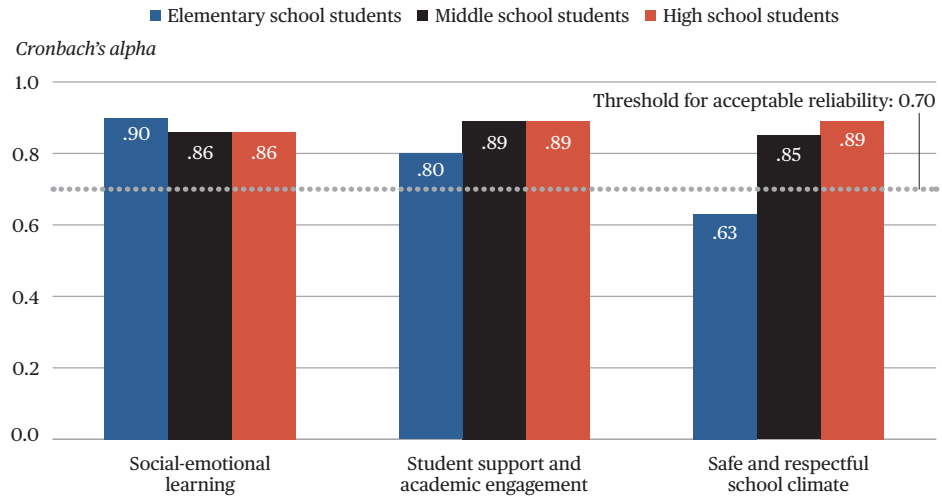
Because the safe and respectful school climate domain for elementary school students did not meet the threshold for reliability, the rest of the report does not present results for the safe and respectful school climate domain or the school climate index for elementary school students. For this reason many of the remaining findings focus on school climate scores in middle and high schools.

Approaches to scoring the school climate survey in individual schools

This section describes findings to help guide PDE's processes for combining survey responses into school-level scores and providing score reports to schools.

On average within schools, noninstructional staff and teachers had more positive perceptions of school climate than students. The extent to which perceptions of school climate varied by respondent type within schools helps

Figure 1. The safe and respectful school climate domain of the Pennsylvania School Climate Survey for elementary school students fell below the threshold for acceptable reliability, 2021/22



Note: Elementary school student estimates are based on responses to 16 social-emotional learning items, 10 student support and academic engagement items, and 4 safe and respectful school climate items in 127 elementary schools. Middle school student estimates are based on responses to 11 social-emotional learning items, 16 student support and academic engagement items, and 10 safe and respectful school climate items in 70 middle schools. High school student estimates are based on responses to 11 social-emotional learning items, 15 student support and academic engagement items, and 10 safe and respectful school climate items in 73 high schools.

Source: Analysis of data from the Pennsylvania School Climate Survey, 2021/22 school year.

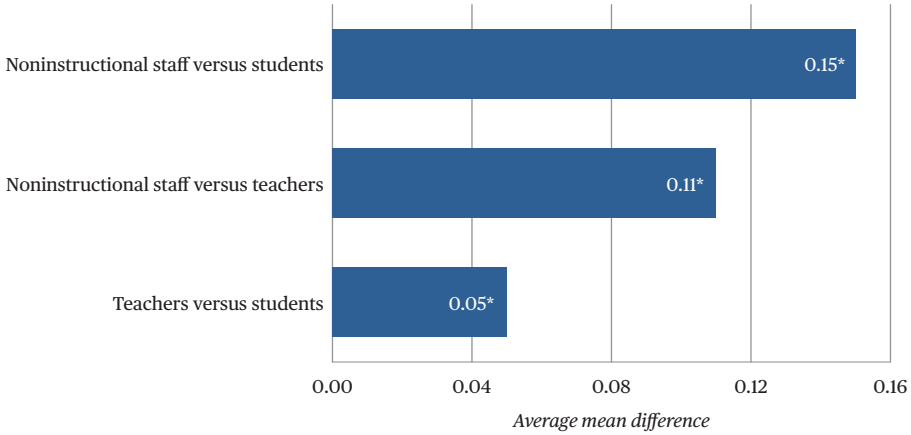
in understanding the extent to which the approach used to weight scores by respondent type within schools might shift school climate summary index scores. On average, noninstructional staff scores were 0.15 point higher than students' scores and 0.11 point higher than teachers' scores on the four-point survey scale ($p < .001$ for both comparisons; figure 2; see also table B20 in appendix B). Both differences represent large effect sizes of 0.74 (between noninstructional staff and students) and 0.42 (between noninstructional staff and teachers). On average, teachers' scores were also higher than students' scores, with a difference of 0.05 on the survey scale, corresponding to a substantive effect size of 0.24 ($p = .04$).

Although perceptions of school climate among the three respondent types differed, correlations were positive between all three types' scores. Student scores were strongly correlated with teachers' scores ($r = .56$; see table B24 in appendix B) and moderately correlated with noninstructional staff's scores ($r = .47$; see table B25), and teachers' scores were strongly correlated with noninstructional staff's scores ($r = .79$; see table B26).

State-weighted scores tended to be lower and have less variation than group-weighted scores. In middle schools and high schools state-weighted school climate summary index scores tended to be lower than group-weighted scores (figure 3). This is expected because students' scores tended to be lower than teachers' and noninstructional staff's scores and are given more weight in the state-weighted approach (.87 in the state-weighted approach and .50 in the group-weighted approach). For middle schools and high schools the group-weighted average school climate summary index score was 2.75 (standard deviation = 0.17), and the state-weighted average was 2.71 (standard deviation = 0.14; see table B27 in appendix B).

In addition, state-weighted scores had less variation than group-weighted scores, resulting in more scores close to the average (see figure 3). This is also expected because teachers' and noninstructional staff's scores had more variation than students' scores and are weighted less in the population-weighted approach. The greater variation in scores may be an advantage of the group-weighted scores–because these scores may better highlight

Figure 2. Within-school mean differences in school climate index scores on the Pennsylvania School Climate Survey are highest between noninstructional staff and students, 2021/22

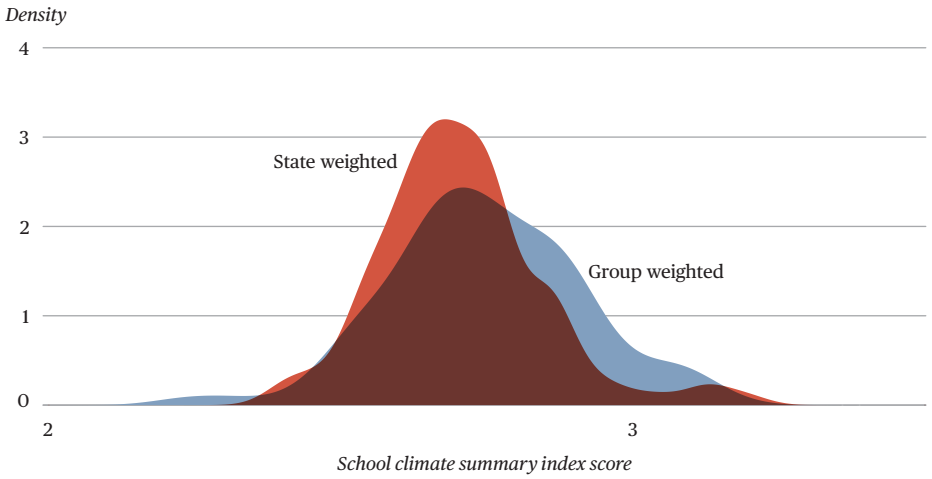


* Indicates a statistically different mean difference at $p < .05$.

Note: Based on data from 79 middle schools and high schools with at least 10 student respondents, 5 teacher respondents, and 5 noninstructional staff respondents. Means for schools that meet at least one of these criteria are reported in tables B16–B19 in appendix B. Average mean differences represent simple averages of survey items that had a scale of 1 (strongly disagree) to 4 (strongly agree). In all domains 1 is the most negative perception of school climate, and 4 is the most positive. The effect size is a calculation of the standardized difference between the groups using Cohen's *d*. Student school climate scores were weighted to account for imbalances between the racial/ethnic and gender composition of respondents and that of their school's population. The difference between the scores of noninstructional staff and students translates to a large effect size of 0.74, the difference between the scores of noninstructional staff and teachers translates to a large effect size of 0.42, and the difference between the scores of teachers and students translates to a substantive effect size of 0.24.

Source: Analysis of data from the Pennsylvania School Climate Survey, 2021/22 school year.

Figure 3. State-weighted summary index scores on the Pennsylvania School Climate Survey for middle and high schools tended to be lower and have less variation than group-weighted scores, 2021/22



Note: Based on scores from 104 schools (42 middle schools and 62 high schools). Elementary schools are not included because the safe and respectful school climate domain did not meet the reliability threshold for elementary school students.

Source: Analysis of data from the Pennsylvania School Climate Survey, 2021/22 school year.

areas for improvement–but state-weighted scores ensure equal representation of each individual's scores, while the group-weighted scores inflate the weights of individual teachers' and noninstructional staff's scores when there are small numbers of those respondents.

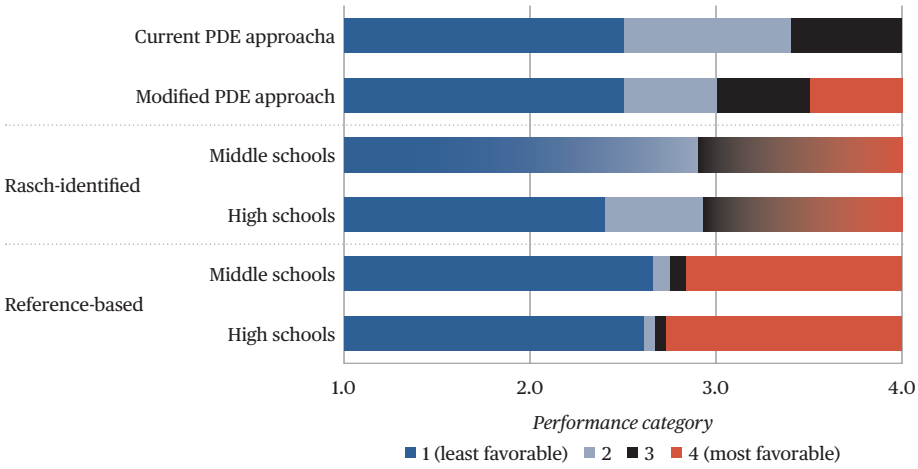
The current PDE approach to defining performance categories resulted in nearly all schools falling into a single category, whereas the other approaches led to more variable distributions. PDE and schools can use performance categories to help identify issues that need to be addressed to improve school climate. The study team used four approaches to identify thresholds for performance categories: the current PDE approach, a modified PDE approach, a Rasch-identified approach, and a reference-based approach (see box 2).

Under the Rasch-identified approach the thresholds between categories 2 and 3 were close to the thresholds between categories 2 and 3 under the modified PDE approach (figure 4). However, under the Rasch-identified approach the thresholds between categories 1 and 2 (for middle schools) and the thresholds between categories 3 and 4 (for middle and high schools) could not be translated to the mean-based survey scale because no school in the sample had a Rasch-scaled score at those thresholds (for example, if the highest school-level Rasch-scaled score in the sample was 63, the Rasch-identified threshold on the mean survey scale corresponding to a Rasch-scaled score of 80 could not be calculated). See table A14 in appendix A for all thresholds using each approach.

Because there was little variation in school climate scores in the sample, the thresholds with population-weighted scores were all much closer together under the reference-based approach than under other approaches.[2] In middle schools the thresholds for the index score were 0.09 point apart on the 1-4 scale, and in high schools the thresholds were 0.06 point apart. By contrast, under the three other approaches the thresholds were 0.5 point or more apart. Therefore, the performance categories from the reference-based approach would

2. To simplify the presentation, the study team focused on state-weighted school climate summary index scores in this finding. The thresholds for the performance categories and distributions of scores for group-weighted scores showed a similar pattern and are reported in table A14 in appendix A.

Figure 4. Thresholds between performance categories on the Pennsylvania School Climate Survey varied according to the approach taken to define categories, 2021/22



PDE is Pennsylvania Department of Education.

Note: The figure shows the range of school climate summary index scores that would be assigned to each performance category based on the approach used. The current and modified PDE approaches assign categories to all school levels using the same predetermined two (for the current PDE approach) or three (for the modified approach) thresholds. The Rasch-identified approach and the reference-based approach define thresholds separately for each school level (middle and high school), and those thresholds are based on population-weighted summary index scores from 42 middle schools and 62 high schools. The index score for elementary school students could not be calculated because the safe and respectful school climate domain did not meet the reliability threshold. Due to very few school climate index scores at the extremes (1 and 4), it was not possible to convert thresholds under the Rausch-identified approach to a score on the mean-based survey scale for the thresholds between 1 and 2 (for middle schools) and the thresholds between 3 and 4 (for both middle and high schools). This phenomenon is shown in the exhibit with a gradient.

a. The current PDE approach has only three performance categories: scores less than or equal to 2.5 are defined as category 1 (least favorable), scores greater than 2.5 and less than 3.4 are defined as category 2, and scores greater than or equal to 3.4 are defined as category 3 (most favorable).

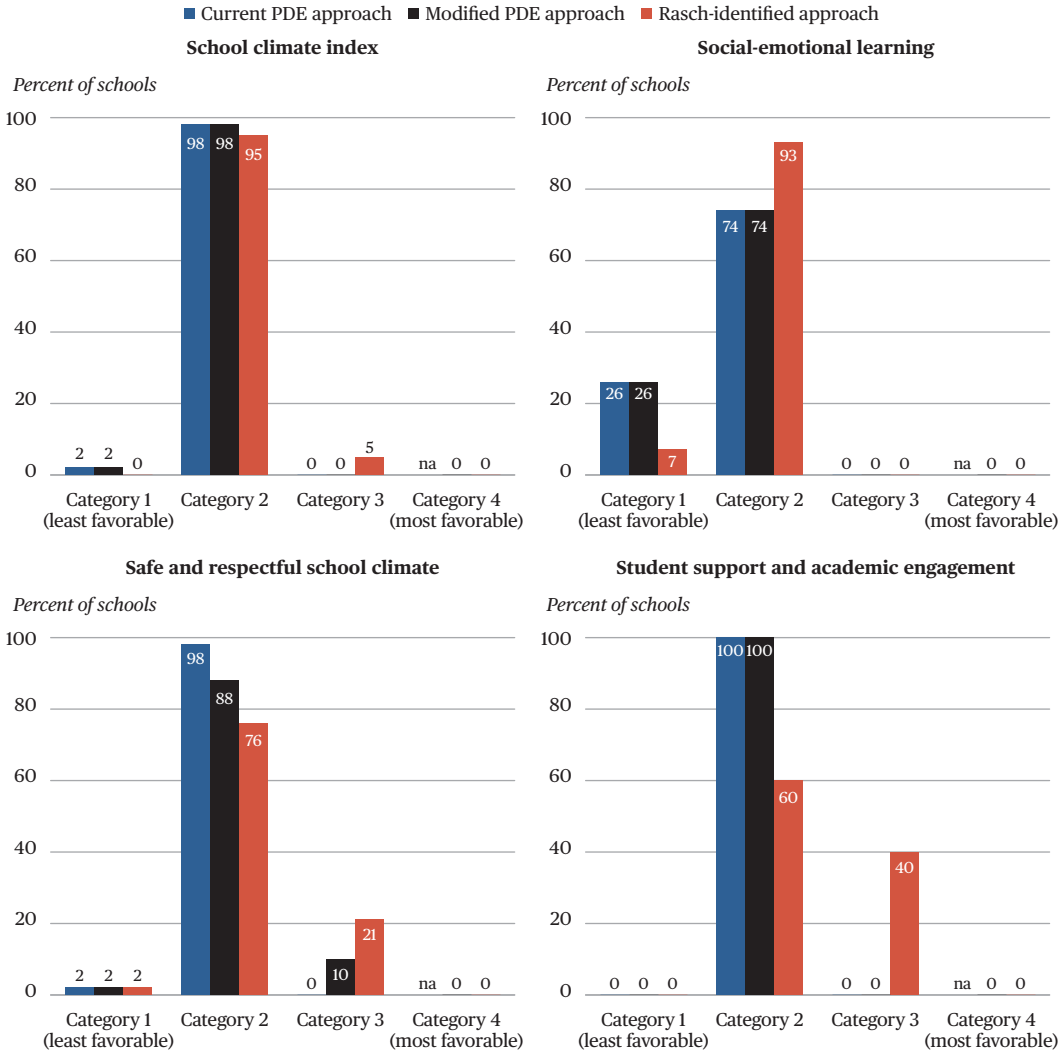
Source: Analysis of data from the Pennsylvania School Climate Survey, 2021/22 school year.

signal smaller differences in school climate that might not be meaningful, depending on the schools included in the sample.

With state-weighted school climate summary index scores, the distribution of schools across performance categories was similar under three approaches to defining performance category thresholds: the current PDE approach, the modified PDE approach, and the Rasch-identified approach (the reference-based approach is not discussed here because 25 percent of schools will always fall into each performance category). Under each of these three approaches, at least 95 percent of middle schools fell into category 2 based on their summary index score (figure 5). Under the current PDE and modified PDE approaches the remaining 2 percent of schools were in category 1, whereas under the Rasch-identified approach the remaining 5 percent of schools were in category 3. Under all three approaches there were no middle schools in category 4. Index scores are not available for elementary schools because the safe and respectful school climate domain did not meet the reliability threshold, but the results for high schools are similar to those for middle schools (see table B38 in appendix B).

The distribution of middle schools across performance categories based on state-weighted domain summary scores varied slightly more under the modified PDE approach and Rasch-identified approach than under the current PDE approach. For example, for the safe and respectful school climate domain, middle schools fell into three performance categories under the modified PDE and Rasch-identified approaches but just two categories under the current PDE approach. Additionally, for the student support and academic engagement domain, middle schools fell into two performance categories under the Rasch-identified approach but just one category under the current and modified PDE approaches. However, for the social emotional learning domain, the

Figure 5. The distribution of middle school climate summary index scores on the Pennsylvania School Climate Survey showed little variation across performance categories regardless of the approach used to define the categories; domain scores showed more variation, 2021/22



na is not applicable because the current PDE approach has only three categories. PDE is Pennsylvania Department of Education.

Note: Based on state-weighted school climate summary scores from 42 middle schools. See table B37 in appendix B for detailed data and tables B36 and B38 for the distributions of schools across performance categories for elementary and high schools.

Source: Analysis of data from the Pennsylvania School Climate Survey, 2021/22 school year.

distribution was similar under all three approaches, with schools falling into just two performance categories. (See tables B36 and B38 in appendix B for the distributions of schools into performance categories for elementary and high schools.)

There is a strong, positive relationship between scaled Rasch scores and mean-based scores for school climate index scores and domain scores for all respondent types and at all school levels. For school climate index and domain scores for each respondent type and for school-level summary scores, the study team found very strong correlations between mean-based scores and scaled Rasch scores ($r > .95$ for all comparisons), demonstrating that mean-based scores and scaled Rasch scores largely convey similar information (see tables B39–B46 in appendix B). All correlations exceeded .90, which suggests that mean-based scores could be used in place of scaled Rasch scores when reporting on school climate.

Perceptions of school climate across student and school groups

This section describes findings related to differences in perceptions of school climate across students with different demographic characteristics within schools and across schools with different characteristics.

Within schools, perceptions of school climate varied by race/ethnicity, gender, and grade level. Three consistent patterns emerged in the comparison of school climate index scores across students of different races/ethnicities in the same school. First, school climate perceptions among the largest racial/ethnic groups (White, Black, and Hispanic students) were similar. The average difference in index score between each of these groups was less than 0.04 (effect sizes < 0.2 and $p > .05$ for all comparisons; figure 6; see also table B47 in appendix B). Second, Asian or Pacific Islander students tended to have more favorable perceptions of school climate, with average index scores 0.08–0.14 higher than those of students of other races/ethnicities in their schools (effect size > 0.3 and $p < .05$ for all comparisons). Third, multiracial students tended to have less favorable perceptions of school climate, with average index scores 0.04–0.14 lower than those of students of other races/ethnicities (effect size < –0.2 and $p < .05$ for all comparisons).

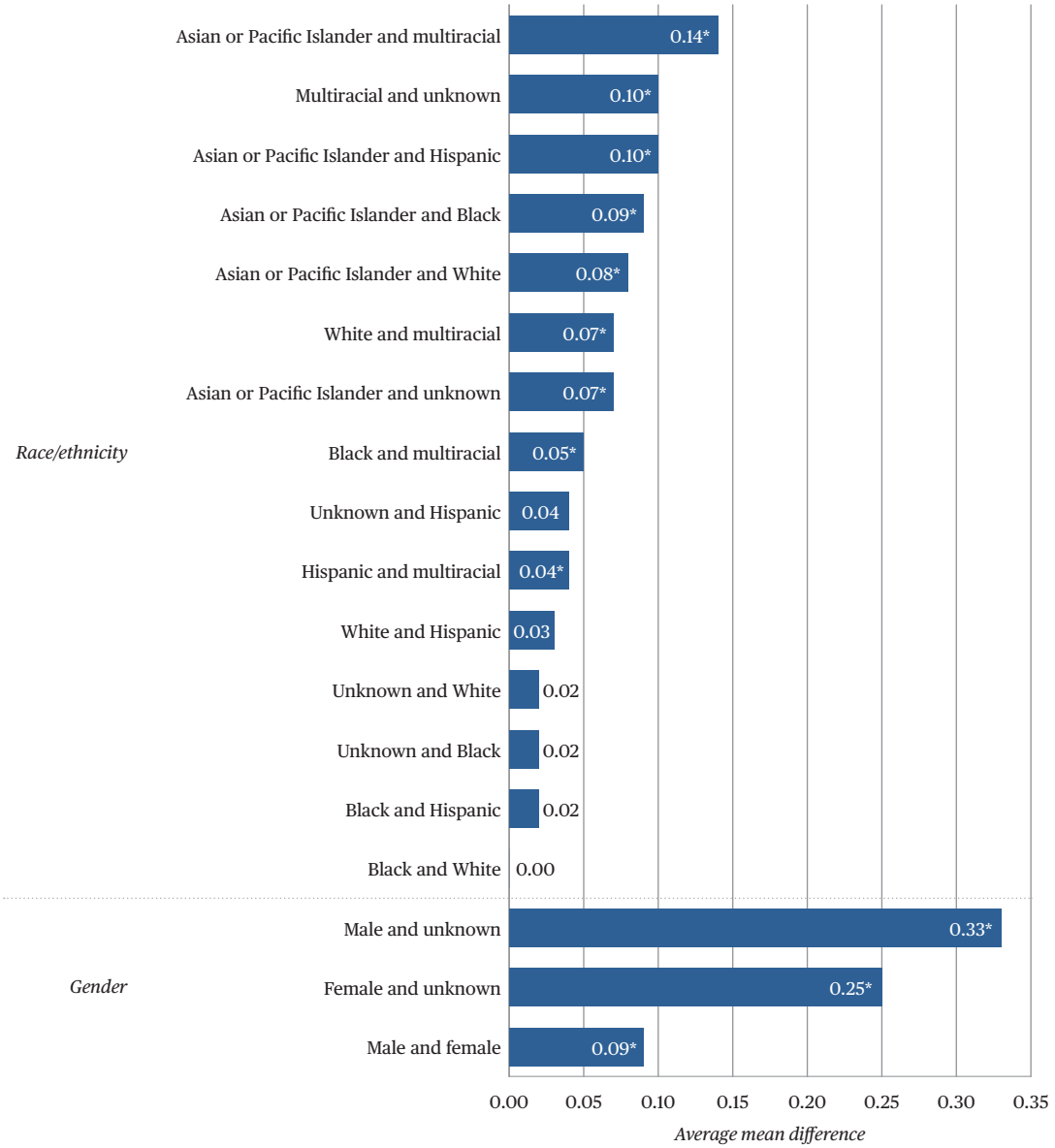
In addition, perceptions of school climate varied by gender. On average, school climate index scores were 0.09 higher among boys than among girls in the same school (effect size = 0.54 and $p < .001$; see figure 6 and table B48 in appendix B). In addition, girls and boys reported more favorable perceptions of school climate compared with students for whom data on gender was unknown, with average index scores 0.25 and 0.33 point above those with unknown gender (effect size > 1.3 and $p < .001$ for both comparisons).

Perceptions of school climate across grade levels showed larger differences in middle and elementary schools (for the available domains) than in high schools. In middle schools, students in lower grades had more favorable perceptions than students in higher grades (effect size > 0.30 and $p < .02$ for all comparisons; see table B50 in appendix B). Similarly, in elementary schools, students in lower grades had more favorable perceptions than students in higher grades for the domains that were scored: social-emotional learning (effect size > 0.34 and $p < .001$ for all comparisons except grade 4 to grade 5) and student support and academic engagement (effect size > 0.54 and $p < .001$ for all comparisons; see table B49). In high schools, perceptions were more similar across all grades (effect size of 0.05–0.23; $p > .28$ for all comparisons; see table B51).

Larger schools and schools with more Black students tended to have lower school climate scores than other schools. The relationship between school climate summary index scores and school size was moderate and negative: schools with more students tended to have lower average school climate index scores (correlation coefficient = –0.44, $p < .001$; table 1). This relationship is negative even when school level was accounted for. Across all available index scores and domain scores and at all school levels, school size had low to moderate correlations with school climate (see table B53 in appendix B). This relationship was strongest in high schools, with moderate correlations for the index and each domain.

Additionally, there was a moderate negative relationship between school climate summary index scores and the percentage of Black students in a school: schools with a larger proportion of Black students tended to have lower average scores (correlation coefficient = –.36, $p < .001$; see table 1 and table B52 in appendix B). This likely reflects the previously documented finding that Black students are more likely to attend schools with more disciplinary incidents, less experienced teachers, and larger class sizes—all school characteristics that are associated with worse school climate ratings (Graham 2022).

Figure 6. Within schools, students' school climate index scores on the Pennsylvania School Climate Survey varied by race/ethnicity and gender, 2021/22



* Indicates a statistically different difference at $p < .05$. All *t*-tests were conducted with a false discovery rate correction to account for multiple comparisons.

Note: Based on scores from 128 middle and high schools. Excludes elementary schools, for which school climate index scores could not be calculated because the safe and respectful school climate domain did not meet the threshold for reliability. Mean school climate index scores are simple averages of survey items that had a scale of 1 (strongly disagree) to 4 (strongly agree). In all domains 1 is the most negative perception of school climate, and 4 is the most positive. Unknown race/ethnicity refers to students who selected "unavailable/unknown/decline." Unknown gender refers to students who selected "not listed." The following differences translate to a large effect size (0.30 and above): Asian or Pacific Islander and multiracial, multiracial and unknown, Asian or Pacific Islander and Hispanic, Asian or Pacific Islander and Black, Asian or Pacific Islander and White, White and multiracial, Asian or Pacific Islander and unknown, male and unknown, female and unknown, and male and female. The following differences translate to a substantive effect size (0.20–0.29): Black and multiracial, unknown and Hispanic, and Hispanic and multiracial. The following differences translate to a moderate effect size (0.10–0.19): White and Hispanic, and unknown and White. The following differences translate to a small effect size (0.00–0.09): Unknown and Black, Black and Hispanic, and Black and White.

Source: Analysis of data from the Pennsylvania School Climate Survey, 2021/22 school year.

Table 1. Correlations between state-weighted school climate index scores on the Pennsylvania School Climate Survey and school characteristics, 2021/22

School characteristic	Correlation coefficient[a]	*p*-value
School size	-.44*	< .001
Percentage of Asian or Pacific Islander students	-.01	.903
Percentage of Black students	-.36*	< .001
Percentage of Hispanic students	-.11	.274
Percentage of multiracial students	-.22*	.027
Percentage of White students	.22*	.027
Percentage of students eligible for the National School Lunch Program	-.13	.186
Percentage of students proficient in math	-.02	.872
Percentage of students proficient in reading/language arts	.19	.058

* Indicates a correlation value that is statistically different from 0 at $p < .05$.

Note: Based on scores from 128 middle and high schools. Excludes elementary schools, for which school climate index scores could not be calculated because the safe and respectful school climate domain did not meet the threshold for reliability. Mean school climate scores are simple averages of survey items that had a scale of 1 (strongly disagree) to 4 (strongly agree). In all domains 1 is the most negative perception of school climate, and 4 is the most positive. Student school climate scores were weighted to account for imbalances between the racial/ethnic and gender compositions of respondents and those of their school's population.

a. The correlation coefficient is reported as Pearson's *r* and can be interpreted as follows: below .30 is low, .30–.49 is moderate, and .50 and above is strong.

Source: Analysis of data from the Pennsylvania School Climate Survey, 2021/22 school year.

Implications

To inform PDE's future efforts to make the Pennsylvania School Climate Survey more useful and actionable, the study team identified several key decision points for consideration.

Bolster the safe and respectful school climate domain of the elementary school student survey

Because two items in the safe and respectful school climate domain of the elementary school student survey did not meet the construct validity criteria, PDE should consider removing those items from future survey administrations. In addition, since the safe and respectful school climate domain did not meet the reliability threshold, PDE should consider adding items to this domain from surveys previously validated for elementary school students. The lower reliability estimate for the safe and respectful school climate domain and the borderline discriminant validity between this domain and the student support and academic engagement domain suggest that the elementary school student survey would benefit from revisions to the safe and respectful school climate domain. After items that did not meet construct validity criteria were removed from the safe and respectful school climate domain, four items remain. In comparison, the middle school student survey includes 10 items for this domain, and the high school student survey includes 16 items. PDE could replace the small number of items in the safe and respectful school climate domain of the elementary school student survey with an existing scale measuring a similar construct on another instrument produced by another organization,[3] if the validity and reliability of the other instrument have been tested and confirmed for elementary school students. PDE has previously had success using this approach to strengthen the social-emotional learning domain to address low reliability for elementary school students.

3. One promising option is the Delaware School Survey, which includes three subscales in the Delaware School Climate Scale (tested for validity and reliability on a population of students from grades 3-12) aligned with the topics currently covered in PDE's safe and respectful school climate domain: student-student relations, school safety, and bullying schoolwide (Bear et al., 2021).

Determine an approach for scoring the school climate survey that meets the Pennslvannia Department of Education's needs

When comparing two approaches to weighting students and staff within schools, PDE will need to decide whether to have equal representation of staff and students in the index scores (group weighted) or to use an approach that reflects the proportions of each respondent in the state, giving more weight to student scores (state weighted). PDE may prefer the state-weighted approach if it wants each student, teacher, and noninstructional staff person to have implicitly equal representation in the survey. An advantage of the state-weighted approach is that it may produce more consistent results than the group-weighted approach because it does not inflate the weights of individual teachers' and noninstructional staff's scores when there are small numbers of those respondents. However, an advantage of the group-weighted approach is that it shows greater variation in scores across schools, because teachers and noninstructional staff, who have greater variation in responses than students, have greater weight. Using the group-weighted approach could lead to more variation in the distribution of schools across performance categories, which could help schools identify areas in need of improvement.

On whether to use mean-based scores or Rasch scores, the very strong, positive relationships between the two types of scores suggest that they are relaying similar information and that PDE would be justified in using either approach. PDE currently uses mean scores in its score reports and has existing infrastructure for analyzing school climate scores in this way, so PDE might find it more efficient to continue using mean scores rather than switching to Rasch scores. In addition, mean-based scores are more intuitive and transparent in terms of calculating and interpreting than Rasch scores. PDE might wish to maintain the current scoring approach given that it is simpler for PDE to calculate and for schools to understand yet provides comparable information as the Rasch approach.

Because the current approach to performance thresholds results in nearly all schools being sorted into the same performance category, PDE should consider using other thresholds that spread the distribution across more categories and providing additional information about true differences in performance. The current approach has the advantage of being the current known method for schools that have received reports up to this point and is derived directly from the survey scale. The modified PDE approach has the same benefit of intuitive interpretation, as it is derived directly from the survey scale. It also slightly improves the variability of schools across performance categories by splitting the middle category into two.

The Rasch-identified thresholds have the benefit of being data-driven, as they estimate the point at which respondents switch from being more likely to select one response category to being more likely to select the next response category. However, these thresholds are more cumbersome to calculate, and because the thresholds have been translated from Rasch units to mean score units, they are not related to the natural units of the survey scale. For example, a threshold could be at 3.64, rather than at a more natural threshold given the 1–4 scale. In addition, given how dependent Rasch-identified thresholds are on the sample of respondents, the thresholds that the study team calculated for 2021/22 might not fit for previous or future years. This could lead to difficulty in understanding trends, given shifting thresholds year to year, and to misleading interpretations of trends.

The reference-based approach, which uses quartile scores of all schools that took the survey as thresholds, has the benefit of being easy to interpret and providing individual schools with information about how their own school climate compares with that of other schools. However, under this approach, performance categories are heavily influenced by the schools in the sample—which is likely to vary year to year. Like Rasch-identified thresholds, reference-based thresholds may be difficult to use for interpreting trends over time. If a school moves up from category 2 to category 3, for example, the school might assume that its school climate improved when the change was actually due to different schools being in the sample. Because schools might still be interested

in understanding their performance relative to other participating schools, PDE could consider giving schools their percentile rank while using another approach to define performance categories. Percentile ranks would need to come with a caveat that schools in the sample are not representative of schools in the state.

Table 2 summarizes the considerations for defining performance thresholds and the approaches that are most and least favorable for each consideration.

Table 2. Considerations for approaches to defining performance categories for the Pennsylvania School Climate Survey

Factor	Stronger approaches	More limited approaches
Ease of interpretation for schools using the score reports	**Current and modified PDE:** Thresholds between performance categories are aligned with scale.	**Rasch-identified:** Thresholds between performance categories are less intuitive and may change from year to year and across domains.
Ease of calculation for producing the score reports	**Current and modified PDE:** Requires comparing scores to a threshold that is constant from year to year and across survey domains.	**Rasch-identified:** Requires running Rasch models, which calls for some psychometric expertise and specialized software.
Sensitivity to true differences between schools	**Rasch-identified:** Estimates the point at which respondents switch from being more likely to select one response category to being more likely to select the next response category.	**Reference-based:** Based on performance of other participating schools.
Stability of performance thresholds over time	**Current and modified PDE:** Thresholds do not shift from one year to the next.	**Rasch-identified and reference-based:** Thresholds change each year, making it difficult to track trends over time.
Distribution of schools into multiple, reliably distinguishable performance categories	**Modified PDE and Rasch-identified:** Schools are sorted into more performance categories than under the current PDE approach, and categories are reliably distinguishable.	**Current PDE:** Nearly all schools are in the same performance category.

Source: Authors' compilation.

References

Aldridge, J. M., & McChesney, K. (2018). The relationships between school climate and adolescent mental health and wellbeing: A systematic literature review. *International Journal of Educational Research, 88,* 121-145. https://www.sciencedirect.com/science/article/pii/S0883035517313794

Amos, L., & Xue, Y. (2021). *Development of Pennsylvania Department of Education School Climate Index summary*. Regional Educational Laboratory Mid-Atlantic. https://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_MA_5.2.5_School_Climate_memo_508c.pdf

Bear, G., Yang, C., Harris, A., Mantz, L., Hearn, S., & Boyer, D. (2021). *Technical manual for Delaware School Survey: Scales of School Climate; Bullying Victimization; Student Engagement; Positive, Punitive, and Social Emotional Learning Techniques; and Social and Emotional Competencies*. Delaware Positive Behavior Support and School Climate & Student Success Projects.

Bentler, P. M. (1990). Fit indexes, Lagrange multipliers, constraint changes and incomplete data in structural models. *Multivariate Behavioral Research, 25*(2), 163-172. https://www.ncbi.nlm.nih.gov/pubmed/26794478

Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *BMJ, 314*(7080), Article 572. https://www.ncbi.nlm.nih.gov/pubmed/9055718

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

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

Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record, 111*(1), 180–213.

Collie, R. J., Shapka, J. D., & Perry, N. E. (2012). School climate and social–emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology, 104*(4), 1189. https://doi.org/10.1037/a0029356

Daily, S. M., Mann, M. J., Kristjansson, A. L., Smith, M. L., & Zullig, K. J. (2019). School climate and academic achievement in middle and high school students. *Journal of School Health, 89*(3), 173–180.

Dogan, E. (2018). An application of the partial credit IRT model in identifying benchmarks for polytomous rating scale instruments. *Practical Assessment, Research, and Evaluation, 23*(7). https://eric.ed.gov/?id=EJ1180135

Graham, J. (2022). Explaining the racial school climate gap: Evidence from Georgia. *AERA Open, 8*. https://journals.sagepub.com/doi/10.1177/23328584221131529

Grayson, J. L., & Alvarez, H. K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and Teacher Education, 24*(5), 1349–1363. https://doi.org/10.1016/j.tate.2007.06.005

Kautz, T., Feeney, K., Chiang, H., Lauffer, S., Bartlett, M., & Tilley, C. (2021). *Using a survey of social and emotional learning and school climate to inform decisionmaking.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. https://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2021114REV.pdf

Kwong, D., & Davis, J. R. (2015). School climate for academic success: A multilevel analysis of school climate and student outcomes. *Journal of Research in Education, 25*(2), 68–81.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

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.* https://schoolclimate.org/wp-content/uploads/2021/05/school-climate-challenge-web.pdf

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual spring meeting of the Psychometric Society, Iowa City, IA.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10. https://eric.ed.gov/?id=EJ075116

Voight, A., & Hanson, T. (2017). *How are middle school climate and academic performance related across schools and over time?* (REL 2017–212). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. https://ies.ed.gov/ncee/edlabs/regions/west/pdf/REL_2017212.pdf