

JONATHAN SCHWEIG, LAURA S. HAMILTON, GARRETT BAKER

School and Classroom Climate Measures

Considerations for Use by State and Local Education Leaders

Research confirms the widespread belief that student learning is influenced by features of the school and classroom environments in which instruction takes place. These qualities of the learning environment, often referred to as *school and classroom climate* (a concept we define in detail later), are associated with higher student achievement (Allensworth, Farrington, et al., 2018; Durlak et al., 2011; Shindler et al., 2016; Wang and Degol, 2016; Aspen Institute, 2019), improved attendance and graduation rates, and lower rates of suspension (Christle, Jolivette, and Nelson, 2007). Studies also indicate significant associations between positive climates and school engagement, motivation, and self-efficacy (e.g., Fast et al., 2010). Although most of this research examines correlations and does not provide rigorous evidence of a causal relationship between climate and other outcomes, the Centers for Disease Control and Prevention (2009) and the Institute of Education Sciences (Dynarski et al., 2008) have identified school climate improvement as an evidence-based strategy that can reduce dropout rates and promote students' feelings of connectedness to school (Thapa et al., 2013).

The evidence of climate's importance continues to accrue. At the same time, educators' appreciation of the value of creating positive, safe, and inclusive school and classroom climate has grown steadily, and educators and policymakers are increasingly making school climate improvement a focus of their work (Hamilton, Doss, and Steiner, 2019). Concurrently, states, districts, and schools have begun to measure climate in order to monitor it and incentivize improvements. At the federal level, the Every Student Succeeds Act (ESSA) permits states to include school climate as an indicator of school quality as part of their accountability and improvement systems (Hough, Penner, and Witte, 2016; Kostyo, Cardichon, and Darling-Hammond, 2018). Eight states are using student climate surveys in school accountability systems, and an additional six states have articulated strategies to make school

climate data publicly available, although these data are not being used for accountability or improvement purposes (Kostyo, Cardichon, and Darling-Hammond, 2018). California’s CORE Districts have implemented a school accountability system that includes measures of school climate derived from surveys of parents, teachers, and students (Marsh, Bush-Mecenas, et al., 2016).

Climate measures are widely used to support decisions at the school and classroom levels about programs and practices (Hamilton, Doss, and Steiner, 2019). School climate indicators also have been incorporated into principal evaluation systems across the United States (Clifford et al., 2012). According to the Center on Great Teachers and Leaders (undated), 22 states recommend or require that school climate surveys be represented in principal evaluation, and an additional five states allow school climate surveys as one possible component of a comprehensive evaluation model.

To support strategies that are focused on creating positive, safe, and inclusive school and classroom climate, educators need to be able to define the specific features of the learning environment on which to focus. There are many different, interrelated aspects of school and classroom climate, and it is important for stakeholders to have a clear sense of what is meant, exactly, by the term *climate*. Educators also need to understand how features of school and classroom climate can be operationalized so that they are measurable. In other words, educators need answers to at least two questions:

- What is meant by *climate*, exactly?
- How do we measure it?

However, educators often lack access to information that could aid in formulating answers to these questions. In this report, we hope to provide information that will define school and classroom climate, illustrate the importance of assessing climate quality, and assist educators in selecting appropriate climate assessments.

This report has six main sections. In the first section, we define *school and classroom climate* and describe how climate differs from similar concepts, such as culture and context. In the second section, we discuss why stakeholders should consider assessing school and classroom climate. In the third section,

we discuss the strengths and weaknesses of various methods that are conventionally used to assess school and classroom climate, such as surveys and structured observations. In the fourth section, we describe important considerations for climate assessment, such as the importance of attending to subgroup differences and the appropriateness of climate assessments for high-stakes use; in the fifth section, we discuss technical properties of climate assessments, including the extent to which such assessments provide information that is fair, accurate, and precise. At the end of the report, we present a list of resources intended to provide educators with additional information about school and classroom climate and the variety of instruments available for climate assessment.

School and Classroom Climate

School Climate, Defined

School climate is complex, encompassing many different aspects of the school environment (Cohen, McCabe, et al., 2009). Based on conceptions of climate from organizational psychology, *school climate* refers to the qualities of the school environment that are experienced by teachers, administrators, students, and other members of the school community. It is the feel of the school that emerges from the perceptions of these individuals as they experience school policies, practices, and procedures (e.g., Ostroff, Kinicki, and Tamkins, 2003; Hoy, 1990). Because climate is shaped by attributes that are structural (such as the availability of supplies, resources, and materials) and systemic (such as professional development programs and teaching practices) that are built up over a period of time, it is relatively stable and persistent.

In fact, because climate is perceived by individuals, there can be as many school climates as there are members of the school community. People will often react in different ways to the same school policies, and a policy that seems fair to one student might seem unfair to another. When individuals share similar perceptions of the school environment, climate can be described and defined at a system level: School climate emerges as the collective product of individual perceptions (Gray, 2007).

Using School Climate Surveys for Continuous Improvement and Accountability: The Case of CORE

California's CORE Districts serve more than 1 million students in eight large urban centers across the state. Working initially under a No Child Left Behind waiver, the CORE Districts collaborated on the development of a school quality improvement system that incorporates information about academic performance and about social-emotional, culture, and climate factors. Academic measures make up 60 percent of a school's quality improvement score; nonacademic measures make up the other 40 percent. School climate, which is measured through surveys of parents, teachers, and students (Marsh, Bush-Mecenas, et al., 2016), accounts for 8 percent of a school's overall score (West et al., 2017; Toch and Miller, 2019).

Students in grades 4 to 12; teachers and staff; and parents, guardians, and caregivers annually participate in surveys to assess their perceptions of school culture and climate. The surveys capture information about the perceived quality of teaching and learning, interpersonal relationships, safety, and school-community engagement. The items are adapted from the California Healthy Kids Survey, originally created by WestEd for the California Department of Education (Hanson and Kim, 2007). The surveys feature prompts for students to indicate

the extent to which they agree with such statements as:

- I am happy to be at this school.
- I feel like I am part of this school.
- The teachers at this school treat students fairly.

California's CORE Districts used information from the CORE Student Culture and Climate Survey to identify race-based gaps in school climate, develop district-level strategic plans, and implement interventions intended to address educational inequalities (Marsh, Bush-Mecenas, et al., 2016). Leadership teams linked student achievement data with school climate data to investigate the extent to which students on the whole felt supported and engaged by their teachers, and to identify schools where African American and Latino students were either thriving or struggling. This information allowed district leaders to think critically about their strategic plans and to support the rollout and implementation of practices and interventions to address problems where warranted (Nayfack et al., 2017).

To learn more, see the Policy Analysis for California Education's publications webpage (undated) and the case study of the CORE districts (Nayfack et al., 2017).

The U.S. Department of Education (undated) has defined a school climate model that has three distinct characteristics—safety, engagement, and environment—and these characteristics largely align with other categorizations, though it is also common to see the learning or academic environment distinguished from the institutional environment (e.g., Wang and Degol, 2016; Thapa et al., 2013; Kostyo, Cardichon, and Darling-Hammond, 2018). These characteristics are defined as follows:

- *Safety* refers to physical safety (e.g., reduced violence and aggression), social and emotional safety (e.g., lack of instances of bullying or cyberbullying), and the fairness and consistency of school rules. *Safety* can also refer to emergency readiness planning.

- *Engagement* refers to the quality of interactions and relationships among individuals in the school community, addressing such dimensions as school connectedness (e.g., sense of belonging or attachment to the school), trust, respect for diversity and cultural awareness, and leadership (e.g., principal and administrator support for teachers).
- *Academic environment* refers to the quality of instruction, teacher expectations (e.g., cognitive demand of academic tasks, emphasis on student improvement and progress), the use of responsive and supportive teaching practices, and opportunities for teachers to participate in professional development.

Using Classroom Climate Surveys in State Teacher Accountability Systems: The Case of Hawaii

Hawaii's State Department of Education serves nearly 180,000 students in 255 public schools across the state. Hawaii is the only state with a single statewide school system for kindergarten through grade 12 (K–12) that functions as both a state education agency and a local education agency. In 2011–2012, Hawaii developed and piloted the Educator Effectiveness System under a Race to the Top Grant. The Educator Effectiveness System is composed of indicators of teacher practice and student learning and growth. These indicators are derived from multiple measures, such as classroom observations, working portfolios, classroom climate surveys, and achievement growth scores. The system was fully implemented in 2013–2014.

In Hawaii, students in grades 3 to 12 annually participate in surveys to assess their perceptions of classroom climate. From 2011 to 2019, the Educator Effectiveness System used the Tripod Survey (Ferguson, 2010) as its primary instrument. The Tripod Survey assesses seven aspects of climate:

- **Care:** How well does the teacher make students feel cared for? (“My teacher in this class makes me feel that s/he really cares about me.”)
- **Clarify:** How well does the teacher diagnose misunderstanding and use multiple strategies to ensure student comprehension? (“My teacher has several good ways to explain each topic that we cover in this class.”)
- **Consolidate:** How well does the teacher ensure conceptual understanding of course material? (“My teacher takes the time to summarize what we learn each day.”)
- **Captivate:** How well does the teacher maintain students' attention during class and cultivate an engaging instructional climate? (“My teacher makes learning enjoyable.”)
- **Confer:** How well does the teacher encourage student participation and active engagement in the classroom? (“Students get to decide how activities are done in this class.”)
- **Challenge:** How well does the teacher cultivate a learning environment with high academic standards? (“My teacher wants us to use our thinking skills, not just memorize things.”)
- **Control:** How well does the teacher manage classroom behavior? (“Student behavior in this class is a problem.”)

Schools in Hawaii used the Tripod Survey results both to assess teacher accountability and to provide formative feedback to teachers, so that they could focus on improving specific aspects of their instructional practices. Some teachers, for example, used survey results as the basis for written reflections, with a focus on developing strategies and action plans to improve classroom learning environments. Tripod Survey data were also aggregated to the school level, and the data were used to highlight school and district priorities, track improvement, and evaluate educational programs and interventions.

To learn more, see the Tripod homepage (undated).

- *Institutional environment* refers to the availability of resources and the quality of the physical environment, such as the adequacy of technology and instructional materials (including textbooks), building maintenance and quality, and class size.

Classroom Climate, Defined

Whereas *school climate* refers to aspects of the overall school environment, *classroom climate* refers to the

specific instructional environments cultivated by individual teachers. In fact, most students' experiences in schools are situated in specific classrooms and are shaped by classroom-based interactions with peers and teachers. Students could have different perceptions of different classrooms, and each classroom might have a distinct climate that emerges as students interact with each other and with teachers. Classroom climates can vary greatly from classroom to classroom within a school (Creemers, 1994; Wang, Haertel, and Walberg, 1993).

Classroom climate and school climate share many of the same aspects of safety, engagement, academic environment, and institutional environment. Although there is significant overlap between characteristics of school and classroom climate, classroom climate is distinct from school climate in a few key ways. For example, although some aspects of safety, engagement, and environment focus on common areas and public spaces—such as cafeterias, hallways, and libraries—or are subject to schoolwide policies that are generally stable across classrooms (Wang and Degol, 2016), other aspects of the learning environment are likely to vary from one classroom to another, such as aspects of engagement (including student-teacher trust), aspects of the academic environment (including the quality of instruction), and teacher expectations. Importantly, in the United States, most students in secondary schools experience multiple instructional environments within a given day and have interactions with multiple teachers. Here, we list the four climate characteristics defined in the previous section and give some specific examples of how they manifest in specific classrooms:

- *Safety* refers, in addition to physical safety, to the extent to which individuals are empathetic and demonstrate caring and support, the intrapersonal competencies of students and teachers, and the fairness and consistency of classroom rules.
- *Engagement* refers to the quality of student-student relationships and student-teacher relationships.
- *Academic environment* refers to learning expectations, the variety of skills and knowledge that students are expected to demonstrate, and instructional rigor and class content (e.g., mathematics or English Language Arts).
- *Institutional environment* refers to the availability of resources (such as the adequacy of technology and textbooks) and to class size and the physical arrangement of the classroom (e.g., desks in rows, tables organized into workstations).

Although there is significant overlap between characteristics of school and classroom climate, classroom climate is distinct from school climate in a few key ways.

Climate, Culture, and Context

Climate is closely related to two other concepts that are often used to describe the character of a learning environment: culture and context. Although these three terms are often used interchangeably, they refer to different (and overlapping) concepts (Van Houtte, 2005).

Culture refers to the rituals, norms, values, beliefs, and assumptions of a school or classroom. Culture is built through relationships over time and provides a classroom or a school with an identity. Culture is also important for establishing standards for appropriate behavior (Stolp and Smith, 1995) and for establishing the rules that govern how teachers and students interact with one another. Thus, culture informs the way that teaching and learning are carried out in a school (Smircich, 1985). For example, norms of collegiality among teachers promote collaborative planning and effective decisionmaking (Deal and Peterson, 2016). Schools can also establish norms for how students should participate in class discussions, or how faculty members should engage and participate in grade-level meetings. Schools also can have established rituals to celebrate community successes and recognize individual contributions (Fullan, 1998; Louis, 1994).

Context refers both to the compositional characteristics of a school or a classroom and to the characteristics of the neighborhood or surrounding area (Clifford et al., 2012). These factors are specific to a time, place, and population (Marsh, Lüdtke, et al., 2012). For example, the gender composition of a classroom and the socioeconomic status of a school's students are contextual characteristics: From year to year, these characteristics change depending on the specific students who are enrolled. Unlike culture and climate, which can be transformed by changing school-level policies and practices, educators typically have less control over context.

Culture and climate shape one another, and aspects of culture can either inhibit or promote a positive climate. However, research indicates there are two reasons to focus on assessing climate. First, it is widely believed that climate is more malleable than culture—changing culture is difficult, and it is easier to change an organization's climate than to change its norms and beliefs, which requires systematic and comprehensive attention (Gruenert, 2008; Sarason, 1982). Second, because climate is based on perceptions of school or classroom experiences, it might be more easily measured than culture (Hoy, 1990).

School and classroom climate is complex and multidimensional, and the way the two climates are linked involves many interrelated features of the learning environment. In the next section, we discuss some of the reasons why educators should consider assessing school and classroom climate. We then provide some guidance on several widely used climate assessment methods.

Reasons to Assess Climate

As was previously described, positive school and classroom climate is associated with numerous desirable outcomes, such as higher academic achievement, improved attendance and graduation, and reduced suspension. Research also suggests that climate is malleable, and that it is possible for principals and teachers to improve school climate (Clifford et al., 2012; Ferguson, 2010; Balch, 2012; Follman, 1992). To engage in effective climate improvement efforts, educators need useful and timely measures to

monitor progress and ensure that any improvement in climate is reaching all students (Allensworth and Hart, 2018). Although data alone do not always translate into climate improvement or meaningful change (Kyriakides, 2005; Fresko and Nasser, 2001), data about school and classroom climate can provide organizations with the information necessary to diagnose problems and develop interventions that can change classroom or school climate before problems become rooted in school life (MacNeil, Prater, and Busch, 2009; Hoy, 1990; Gruenert, 2008). This research base argues that assessing school and classroom climate allows schools and districts to do the following:

- **Communicate school and classroom climate as a priority.** Committing time and resources to assessing school and classroom climate communicates that creating, supporting, and sustaining positive learning environments is a priority. The decision to assess school and classroom climate can focus stakeholders' attention on the importance of the learning environment, alter their perception of the role of school and classroom climate in overall school improvement, and change their behavior (Gehlbach, Robinson, et al., 2018; Marsh, Bush-Mecenas, et al., 2016).
- **Identify structural problems in schools, which can then be addressed with changes in policy or school-based practice.** Assessments of school and classroom climate can provide diagnostic information that can guide structured opportunities for reflection, discussion, and collaboration among colleagues (Peterson, 2000; Gehlbach, Brinkworth, et al., 2016). For example, climate assessments might (1) yield information about race- or gender-based differences in disciplinary approaches or (2) surface systematic differences in students' opportunities to engage in cognitively demanding learning experiences (Babad, 1993; Brophy and Good, 1974; Knapp, 1995).
- **Provide actionable improvement targets for school personnel.** School or classroom climate assessments can provide teachers and administrators with feedback about areas of

school or classroom life that are functioning well and about areas in need of development (Clifford et al., 2012). This information can assist administrators in developing professional growth plans. For example, climate surveys might help teachers and administrators understand the extent to which students perceive classrooms as emotionally safe, caring, and supportive. In classrooms where students report feeling less supported, teachers and administrators can collaborate on developing and implementing strategies to improve student-teacher relationships. When climate is measured over time using a common approach, the data can provide important information about growth and progress toward school goals.

- **Supplement measures of students' social and emotional skills or competencies.** Assessments of social and emotional learning (SEL) skills or competencies are becoming more widely available to educators, but assessments that focus only on students' skills or competencies can provide limited and potentially misleading information. SEL development, school climate, and classroom climate are interdependent and synergetic. For example, students' relationships with peers and with school staff are critical to the development of SEL skills and competencies (Baker et al., 2008; Hamre and Pianta, 2001; Rimm-Kaufman and Hamre, 2010). Classrooms with warm teacher-student relationships promote positive SEL development (Schonert-Reichl, 2017), and positive peer relationships can, in turn, promote more-positive attitudes about education and school engagement (Epstein, 1983; Phelan, Davidson, and Yu 1998). Using climate measures alongside social and emotional competency assessments can provide a more comprehensive view of the ways that schools and classrooms support the development of these competencies (Taylor et al., 2018).¹
- **Evaluate the effectiveness of programs and interventions.** Districts and schools can use climate assessments to examine the

Assessments of social and emotional learning (SEL) skills or competencies are becoming more widely available to educators, but assessments that focus only on students' skills or competencies can provide limited and potentially misleading information.

effectiveness of programs or interventions that are designed to improve the learning environment. Examples of such programs are restorative justice programs, cultural competency training, responsive professional development and leadership training programs, and increased opportunities for staff collaboration.

- **Support equitable educational outcomes.** Even within the same school or classroom, students might experience climate differently. Prior evidence highlights sizable within-school race-based differences in how students experience discipline, safety, adult-student relationships, and other aspects of the learning environment (Hough, Kalogrides, and Loeb, 2017; Voight et al., 2015). These *climate gaps* (Voight et al., 2015) have potential implications for equity and might indicate that students are provided with differential access to rich learning experiences; some recent research suggests that schools with larger race-based climate gaps

also have larger race-based achievement gaps (Voight et al., 2015).

Methods for Assessing School and Classroom Climate

School and classroom climate is most commonly assessed using one of two methods. By far, the most prevalent method relies on stakeholder perceptions that are captured on survey questionnaires. Surveys typically require respondents to rate their climate perceptions on an ordinal rating scale. Climate measures are then typically created by aggregating these responses to create school- or classroom-level indicators (either by taking simple averages or by computing more-complicated composites that use advanced statistical models). Examples of stakeholders who participate in such surveys are teachers, parents, school leaders, and students. Each stakeholder has a distinct perspective on school and classroom life. The other common method of assessment uses structured observations of classrooms and schools. Each of these methods has distinct limitations. For one thing, all measurement methods contain a certain amount of measurement error, which introduces uncertainty into appraisals of school and classroom climate (e.g., Bell et al., 2012). For another, school systems often struggle to find the time and administrative capacity to collect and analyze climate data and then act on it (e.g., Government Accountability Office, 2013; Rowan, Schilling, et al., 2013; DePaoli, Atwell, and Bridgeland, 2017). Therefore, we provide additional information about advantages and limitations that are more acutely relevant for particular measurement methods.

Student Surveys

Student surveys of school and classroom climate have a long and rich history in both K–12 and higher education settings (e.g., Fraser, 1998; Wang and Degol, 2016). Some surveys direct students to focus on a specific teacher or class; others ask students to consider their school more broadly. Students are qualified to report on school and classroom climate because they experience school-based interactions every day and have extensive knowledge of their teachers (Ferguson,

2012; Follman, 1992). Furthermore, students are able to provide a perspective on school life that might not be captured or observed by other reporters (Downer et al., 2015; Feldlaufer, Midgley, and Eccles, 1988). Asking students about their perceptions of the learning environment acknowledges and legitimizes their experiences and provides an opportunity for students to define school-based issues and have a say in who gets to try to address them (Mitra, 2007). In addition, as noted, student perspectives can be particularly useful for informing efforts to promote equitable experiences and outcomes because students' perspectives often vary within the same school or classroom.

On the other hand, there is some evidence that climate surveys, like all self-report measures, are vulnerable to acquiescence bias (respondents have a tendency to agree with all items), halo effects (positive perceptions about one climate aspect influence perceptions of other aspects), or recall difficulties (respondents might have issues with remembering specific details or events) (see, e.g., Popham, 2013). For example, work by Price (2016) using longitudinal data from Milwaukee's school climate report cards suggests that school-level scores on climate surveys tend to be skewed favorably on all scales. Additionally, climate items might be susceptible to reference bias (students might differ in their implicit standards of comparison) and to cultural bias (even when climate is similar for two students, perceptions of that climate vary depending on students' identities as members of specific subgroups) (Heine et al., 2002; Calarco, 2011; Garza, 2009).

There are also practical trade-offs in terms of feasibility. Student surveys have the benefit of being cost-effective, relatively easy to administer, and feasible to use at scale (e.g., Balch, 2012). However, designing new surveys and evaluating existing surveys to determine the quality of items and scales can be difficult and time-consuming. So can developing administration strategies that maximize thoughtful student participation (Dillman, Smyth, and Christian, 2014; Gehlbach and Artino, 2018). Additionally, surveys require basic literacy skills for respondents to make sense of the items and the judgments they require. In early elementary classrooms, for example, this could make survey administration impractical. What's more, although students are able to report on their

own experiences in classes and schools, they do not have knowledge of the full range of environmental factors that might determine climate, such as curriculum or teacher-administrator relationships (Follman, 1992; Worrell and Kuterbach, 2001; Little, Goe, and Bell, 2009).

Parent and Community Surveys

Although parents and other community members cannot provide the same kind of firsthand knowledge of school and classroom life as students, they can report on the extent to which they are involved and engaged in their students' learning. Parents, guardians, and other community members can also share important information about the extent to which school representatives communicate with parents, the quality of interactions with school representatives, and the extent to which schools engage parents in decisionmaking (Stevens and Sanchez, 1999). Climate surveys administered to parents and community members enable schools and districts to assess the extent to which individuals (and individuals from identifiable subgroups) feel welcome at school, and to conduct outreach activities to build stronger school and community partnerships (Cohen, Pickeral, and McCloskey, 2009).

Like student surveys, parent and community surveys can be subject to various biases related to the self-report nature of these instruments. In addition, it can be challenging to obtain high rates of participation from parents or other community members, and there is a risk that those who do respond will differ in important ways from those who do not. For example, parents who respond to an invitation to complete a survey might be more likely than other parents to have time available to complete the task, or they might be more likely to be fluent in English. Additionally, parents who respond to surveys might be parents who are more engaged and involved in the school. If response rates are low, the inferences that educators make based on the results might be distorted.

School Staff Surveys

School staff—including teachers, instructional support staff, and administrators—can also provide

Although parents and other community members cannot provide the same kind of firsthand knowledge of school and classroom life as students, they can report on the extent to which they are involved and engaged in their students' learning.

valuable input on school climate. Teachers can provide an important perspective on school safety, parent involvement in school, and the quality of instruction. Teachers and other instructional staff are also able to provide information about aspects of school climate that other reporters (such as students) do not have knowledge of, including teacher-teacher relationships and teacher-administrator relationships, information about the extent to which the school supports professional growth, and whether teachers have adequate time and resources for instruction (Hirsch and Emerick, 2006; TNTP, 2012). In addition to all of the limitations previously listed, researchers have expressed concerns about the accuracy of staff surveys and the validity of inferences about classroom climate that are based on these surveys, particularly when teachers are asked to provide information about their own instruction or relationships with students (e.g., Mayer, 1999; Burstein et al., 1995). Research also shows that correlations among indicators obtained from student ratings, teacher

Structured observations typically involve trained raters (e.g., school leaders, instructional coaches, district administrators, or independent external evaluators) who observe instruction and record their observations on a standardized rubric that facilitates credible comparisons across schools or classrooms.

ratings, and observation-based ratings are generally low (Mayer, 1999; Burstein et al., 1995), suggesting that, to a certain extent, the nature of climate is linked to perspective, and that students, teachers, and external observers are describing distinct but interrelated aspects of the school or classroom (Kunter and Baumert, 2006). Low correlations could also result from measurement error that depresses the reliability of scores on these measures; we discuss reliability in more detail next.

Structured Observations

Structured observations typically involve trained raters (e.g., school leaders, instructional coaches, district administrators, or independent external evaluators) who observe instruction and record their observations on a standardized rubric that facilitates

credible comparisons across schools or classrooms (Little, Goe, and Bell, 2009; Pianta and Hamre, 2009). At a school level, structured observations are often included as components of School Quality Reviews, where teams (usually composed of administrators and experienced educators) observe classrooms and other school spaces (such as hallways, libraries, and cafeterias) to collect information about safety, teaching and learning, and the institutional environment (Kostyo, Cardichon, and Darling-Hammond, 2018). Within classrooms, observation protocols can be designed to attend to the quality of the interactions between students and teachers and to the interactions among students themselves (Pianta and Hamre, 2009; Praetorius and Charalambous, 2018). Such interactions are critical elements of a classroom's social and emotional environment (Holahan and Batey, 2019). Classroom observations also can be used to appraise the extent to which teachers are using instructional practices that foster positive social and emotional development, which, as described previously, is highly interrelated with classroom climate (Holahan and Batey, 2019; Yoder, 2014). Structured observations also can offer a perspective on classroom or school climate that is not filtered through stakeholder perception in the same way as survey-based information (Pianta and Hamre, 2009).

Observation-based ratings also have some limitations that users should be aware of. Recent research suggests these ratings can depend on the observer. For example, principals and other school-based personnel tend to rate teachers more favorably than external evaluators (Kraft and Gilmour, 2017; Grissom and Loeb, 2017). There are other weaknesses of structured observations. Observation protocols are often not well equipped to capture within-classroom variability in student experiences, though there is some evidence that student-teacher interactions vary more within classrooms than across them (Croninger and Valli, 2009; Reinholz and Shah, 2018; Cohen and Goldhaber, 2016). Additionally, observation-based methods might be sensitive to classroom composition or teacher characteristics, such as teacher race and student prior achievement (Whitehurst, Chingos, and Lindquist, 2014), and other factors that do not directly reflect instruction or teacher-student interactions, such as grade level

(Cohen and Goldhaber, 2016; Mihaly and McCaffrey, 2014; Steinberg and Garrett, 2016). Obtaining reliable scores from observation rubrics also might pose significant administrative challenges because it is necessary to use data from multiple observers that were collected during multiple lessons (Hill, Charalambous, and Kraft, 2012; Ho and Kane, 2013). Finally, from a practical perspective, observations require significant investment in training, monitoring, administration, and scoring that place practical constraints on the extent to which such approaches are scalable or feasible for use across a large number of classrooms and schools (Balch, 2012).

Other Methods for Climate Assessment

Numerous other methods have been proposed by researchers and practitioners for assessing school and classroom climate, such as conducting interviews or focus groups, which require an interviewer to ask questions or make statements that prompt interviewees to discuss specific topics or issues. Such interviews could be helpful in exploring, for example, the quality of peer relationships or the extent to which students feel supported by teachers and administrators (Freiberg, 1999). As is the case with observations, however, these approaches require a significant investment in training, administration, and analysis, and they might not be feasible at scale.

Administrative data that are routinely collected by schools and districts can serve as proxy measures for school and classroom climate. For example, absenteeism (and chronic absenteeism, frequently defined as a student missing 10 percent or more of school days) has been shown to be associated with school safety (Centers for Disease Control and Prevention, 2009) and can potentially provide information related to school climate (Holahan and Batey, 2019). Information about suspension and expulsion is routinely collected as required by the U.S. Department of Education Office of Civil Rights. When these data are disaggregated by student subgroup, they can be used to describe the extent to which some student groups are disproportionately subjected to exclusionary discipline policies, a key factor in how students perceive school climate (Holahan and Batey, 2019). Because these are proxy

measures and are influenced by factors beyond the school or classroom environment, any inferences about climate that are based on this type of information should be made with caution and ideally in the context of more-direct measures of climate.

Considerations for Climate Assessment

Although assessing school and classroom climate competencies has the potential to provide the benefits previously described—communicating climate as a priority, providing actionable improvement targets, and supporting equitable educational outcomes—the following considerations should be kept in mind:

- **Climate is complex and multidimensional.** Climate is best thought of not as a single characteristic of a learning environment or an organization but as a complex set of inter-related features. As we noted earlier, *climate* can refer to student learning conditions and to teacher working conditions (Hirsch and Emerick, 2006). Although research has begun to coalesce around the features of the learning environment that matter the most (O’Brennan and Bradshaw, 2013; Wang and Degol, 2016), there is no single definition of school and classroom climate. A wide variety of theoretical frameworks—sometimes overlapping, sometimes conflicting—are used by researchers and practitioners (Wang and Degol, 2016). Frameworks can vary substantially in terms of development process, intended uses or purposes, their conceptualizations of climate, and how climate is operationalized for measurement (e.g., Praetorius and Charalambous, 2018). As a result of this complexity, educators can select from a large number of instruments to assess aspects of school and classroom climate (Cohen, Pickeral, and McCloskey, 2009; Freiberg, 1999; Fraser, 1998), but they might need guidance to determine which ones are best suited to their purposes. Several resources can help with these decisions. For example, Freiberg (1999) documented 18 different measures of school climate, including

student, parent, and community surveys (Creemers and Reezigt, 1999; Fraser, 1998; Stevens and Sanchez, 1999), observation and interview protocols (Stringfield, 1999), and measures based on student drawings and journal narratives (Freiberg and Stein, 1999). The National Center on Safe and Supportive Learning Environments (undated) maintains a compendium of school climate assessments that features 48 different surveys that can be administered to students, family, community members, and school-based staff. The Council of Chief State School Officers (CCSSO) recently published a research brief that includes a list of classroom observation protocols that attend to climate (Holahan and Batey, 2019). (In the last section of this report, we present a list of resources related to school climate measurement, including compendiums of measures and resources for educators and schools.)

- **For survey-based climate measures, the appropriate unit of analysis is not always clear.** As described previously, climate is perceived by individuals; in this regard, there can be many different climates within a school or classroom. However, when individuals share perceptions of a school or classroom, it is appropriate to describe climate as a characteristic of the classroom or the school. Educators should pay careful attention to whether it is more appropriate to attend to individual climate perceptions (represented by the survey scores of individual respondents) or to attend to the shared perceptions of individuals within a classroom or school (represented by aggregating the scores of individual respondents to the classroom or school level). In most K–12 applications, climate is implicitly assumed to be a characteristic of the classroom or the school (Marsh, Lüdtke, et al., 2012; Rowan, Raudenbush, and Kang, 1991; Sirotnik, 1980). This is why most schools and school districts use climate measures that are created by aggregating individual survey responses to the school or classroom level. However, there are cases in which it

might be more sensible to focus on individual perceptions. For example, in examining the relationship between classroom climate perception and student motivation, an individual student’s perception of climate might be more important than the shared perception of all students in the class (Lüdtke, Robitzsch, et al., 2009).

- **Many instruments are not developed using best practices in instrument design.** There is extensive literature in survey methods that offers practical guidelines on the importance of question wording, the formulation of response options, and overall instrument organization and formatting. This guidance suggests, among other things, using construct-specific response options instead of agree-disagree anchors, avoiding negatively worded items, and using item formats that are aligned with the questions of interest (Gehlbach, 2015; Gehlbach and Artino, 2018; Dillman, Smyth, and Christian, 2014). In terms of layout and organization, research has shown that each response option should be equally spaced on the page or screen (with the exception of nonsubstantive options, such as “not applicable”), and each response option should be labeled (Artino and Gehlbach, 2012; Krosnick, 1999; Dillman, Smyth, and Christian, 2014). Literature on the development of standardized observation rubrics suggests that it is critical that rubric developers attend to the description of the levels of performance and avoid unclear or subjective language or the need for high inference, carefully selecting the evaluative criteria (or dimensions) of climate to be observed, and carefully writing definitions of these dimensions (Stevens and Levi, 2005). The guidance also recommends that the number of performance levels be appropriate for each evaluative criterion, and that the rubrics not be overly complex—i.e., the number of evaluative criteria that observers are expected to attend to should be limited (Kane and Staiger, 2012; Stevens and Levi, 2005). Developing sound observational rubrics is a complex task (Schoenfeld, 2013). The selection

of evaluative criteria and dimensions of climate is often governed by a specific theory of instruction or social ecology, and these theories might or might not have a strong empirical evidentiary basis (Cohen and Goldhaber, 2016). Additionally, understanding how to articulate performance levels for teaching practices is complex; some research suggests that defining high-quality demonstrations of practices is often difficult and uncertain (Polikoff, 2015; Cohen and Goldhaber, 2016; Stodolsky, 1990).

- **Results from climate surveys can be influenced by the people or organizations to which the items refer.** Many surveys feature items that have both organizational (e.g., “students in this school,” “the teacher”) and individual (e.g., “I,” “me”) referents and that ask students to make judgments about the behavior of others and about their own behavior. Item-referent material merits particular attention; some researchers have raised concerns that item references can have an impact on item responses. Depending on who they are and how they believe survey results will be used, survey respondents might have different motivations for responding to an item that asks them to judge their own behavior rather than the behavior of others. Additionally, there is information asymmetry—a person has deep knowledge of his or her own motivations but must rely on indirect behavioral indicators to judge the motivations of others (Fauth et al., 2019). Other researchers have argued that because climate is defined as a collective or shared experience, it is inappropriate to use items that refer to an individual’s experience (Marsh, Lüdtke, et al., 2012).
- **Culture can influence climate perceptions.** As with all assessments, it is important to examine the extent to which climate assessments might function differently for identifiable subgroups (American Educational Research Association [AERA], American Psychological Association [APA], and National Council of Measurement in Education [NCME], 2014). In terms of

As with all assessments, it is important to examine the extent to which climate assessments might function differently for identifiable subgroups.

surveys, research has found that students in one context might respond differently to items about climate than would students in another context, and perceptions of climate might reflect cultural or contextual factors that cause students to report on climate differently (West et al., 2017; Bankston and Zhou, 2002). For example, Garza (2009) found evidence that Latino high school students perceived caring behaviors differently from their white peers and placed value on different caring behaviors. Calarco (2011) found evidence that help-seeking behaviors differed depending on socioeconomic status and that middle-class students requested more help than their working-class peers—and, as a result, received more assistance and guidance from teachers. In terms of observation protocols, research suggests that observer background can play a key role in the extent to which protocols are used accurately. For example, an observer’s beliefs about the role of the teacher might influence how that observer uses protocols (Cash et al., 2012). Some research has shown that there can be interaction effects between observers and observed subjects, and that ratings can be influenced by matches in gender or socioeconomic status (Moss and Jones, 1977; Gurwitz and Dodge, 1975; Repp et al.,

1988). It is important to understand climate gaps and to rely on climate information to address issues of educational equity within schools, but it is also important to ensure that climate assessment is culturally responsive and invariant across identifiable subgroups, so that gaps in climate perception reflect substantively different experiences within schools rather than mere differences in reporting. Jagers, Rivas-Drake, and Borowski (2018) provide guidance on how to support SEL and climate assessment that is culturally responsive, and they note the importance of assessing both student perceptions and adult competencies related to issues of race and ethnicity and social class. These authors also point out the value of using multiple measures—such as surveys, observations, and interviews—to gather a broad variety of perspectives from students and adults.

- **Overall or aggregate measures of climate can conceal important subgroup differences.** When using survey-based data, climate measures are typically created by aggregating individual responses to create school- or classroom-level indicators. As mentioned previously, this can be done either by taking simple averages or by computing more-complicated composites that use advanced statistical models. In this way, they characterize overall climate, or the typical experience of students or other stakeholders. Standardized classroom observation protocols are often designed to provide information about typical teaching, and they are not structured to appraise differentiation of learning opportunities and variation in instructional practices within classrooms (Reinholz and Shah, 2018; Cohen and Goldhaber, 2016). Even with tools that focus on teacher-student interactions, observers might be asked to attend to what “most” students are doing rather than registering and rating individual teacher-student interactions. Often, observers are not explicitly asked to attend to whether some students are more engaged than others or whether the teacher engages differently with subsets of

students. However, there is often within-classroom or within-school variability in learning opportunities experienced by different students or groups of students, and there are often within-school and within-classroom differences in climate perception that arise depending on students’ identities as members of specific subgroups (Croninger and Valli, 2009; Voight et al., 2015). To the extent that individuals or subgroups of individuals experience climate in meaningfully different ways, reliance on indicators of typical student-teacher interactions or on aggregated survey responses as measures of climate might obscure information about the learning environment (Roberts, Hulin, and Rousseau, 1978) and compromise the validity and utility of these measures to inform continuous improvement or other decisions (Lüdtke, Trautwein, et al., 2006).

- **Respondents’ anonymity on climate surveys creates some opportunities but limits others.** Frequently, a condition of climate survey administration is that the respondents be anonymous. This approach has many advantages: It mitigates data-privacy concerns, potentially eliminates social desirability bias (i.e., respondents giving answers they think others would want to hear rather than truthful responses), and can improve the accuracy of the survey because individuals are less concerned that there will be adverse consequences if they respond negatively (Blad, 2016; Fagell, 2018). However, anonymous administration can also greatly compromise the ability of practitioners and policymakers to understand how climate changes over time because of difficulty in distinguishing between the extent to which climate ratings reflect meaningful change and the extent to which the ratings reflect change in school or classroom composition. For example, anonymous administration often makes it impossible to consider student mobility, which could be particularly salient in high-poverty neighborhoods or districts that serve large shares of at-risk youth.

- **It is difficult to ensure representativeness of climate measurement.** Given a small sample of students, teachers, and lessons, it can be difficult to ensure that specific appraisals of climate can be generalized to the classroom or school as a whole. For example, climate survey response rates can vary widely, and respondents might not be representative of a particular population. Some kinds of respondents, particularly parents, are frequently not responsive to requests to participate in surveys (Nathanson, McCormick, and Kemple, 2013). If individuals who respond to the survey are systematically different in some way from those who do not, this can undermine the credibility of survey-based information and can also compromise the integrity of the collected data. For example, if students receiving high grades were the only ones to respond to classroom climate surveys, or if mathematics teachers were the only ones to respond to school climate surveys, it is unclear whether the climate measures would generalize to the school or classroom as a whole (Rogelberg and Luong, 1998; Rogelberg and Stanton, 2007). Observations employ different kinds of sampling that make it difficult to ensure representativeness. For example, one or two lessons are typically sampled and observed and used as the basis of climate appraisal. However, research has shown that teaching practices can vary greatly from day to day (Rowan and Correnti, 2009), so a small number of observations might not be fully representative of instruction or classroom climate and might not capture rare but important events (Bell et al., 2012).
- **Although some school districts use survey-based climate data for high-stakes decision-making, validity evidence supporting such uses is scarce.** Climate surveys can provide schools and districts with important diagnostic information about instructional quality and the conditions of school life, and they can be used formatively to improve instruction, reallocate resources, or plan interventions to improve interpersonal relationships

[E]mpirical evidence supporting the use of surveys in high-stakes settings (such as school accountability systems or teacher evaluation systems) is scarce, and most instruments have not been validated for use in these contexts.

among students and teachers (Peterson, 2000; Gehlbach, Brinkworth, et al., 2016). Because climate is malleable and directly influenced by teachers and school leaders, climate surveys have been used prolifically in teacher, school, and principal evaluation systems nationwide (Clifford et al., 2012). However, empirical evidence supporting the use of surveys in high-stakes settings (such as school accountability systems or teacher evaluation systems) is scarce, and most instruments have not been validated for use in these contexts (Clifford et al., 2012). At a minimum, use of climate surveys in high-stakes accountability also creates imperatives that schools and districts ensure that participation is adequate, that survey administration conditions are uniform, and that respondents are engaging with the survey in a thoughtful and meaningful way (Gehlbach and Hough, 2018). Additionally, recent guidance from CCSSO suggests that student surveys should be used in teacher and school evaluation systems at most; CCSSO

does not recommend that schools and districts use school staff, parent, or community surveys for these purposes because of the risk of score corruption (Holahan and Batey, 2019). CCSSO also notes that school staff, parent, or community surveys would not meet ESSA requirements that accountability indicators be able to be disaggregated by student subgroup.

Understanding Validity and Reliability

Making sound inferences regarding the quality of school and classroom climate requires instruments that are consistent, accurate, fair, and credible and that actually measure the aspects of climate they purport to measure. To have trust in the data and to make sound decisions, users need evidence that an instrument produces measures that are accurate, consistent, and credible. Such evidence is collectively known as *evidence of validity and reliability*, and it is typically collected by researchers or developers to build an argument that a particular use of an instrument is warranted in a particular context for a particular population (Kane, 2006). Sometimes, publicly available instruments are published with technical documentation that describes the existing validity and reliability evidence, including information on the populations and contexts from which this evidence has been collected. CCSSO recommends that all practitioners considering using climate surveys fully explore the extent to which such evidence is available prior to instrument adoption (Holihan and Batey, 2019). In the remainder of this section, we discuss aspects of reliability and validity that are particularly important in the context of climate measurement. We support the CCSSO recommendation that schools and districts investigate existing reliability and validity evidence when formulating climate assessment plans.

Reliability

Reliability refers to the extent to which an instrument produces scores that are consistent and free of measurement error. It can be thought of as a measure

of the precision of scores. Several factors related to the measure can introduce random error into scores, thereby diminishing their reliability.

This error can stem from several conditions related to the design or administration of an assessment. For example, an assessment of student-teacher relationships might consist of ten items. However, these ten items are not the only items of interest—and plausibly, a different set of items could have been selected and used to appraise the quality of these relationships. In this case, the sampling of items introduces some measurement error into the assessment process—by random chance, some items might have higher ratings than others and other might have lower ratings, so scores are expected to fluctuate randomly depending on the specific items that make up an assessment form. There might be other sources of error in this example, as well—the occasion on which the assessment was administered is also thought of as randomly selected from all possible assessment administration occasions. By random chance, scores might fluctuate from one day to the next. If some of the items on an assessment are free-response items, scores might depend on expert ratings provided by raters who read student responses and rate their quality, completeness, or correctness. Raters are also conceived of as having been selected from a population of all possible raters. By random chance, some raters might be more severe or more lenient than other raters, and this will introduce some random variation in scores that is not related to the construct of interest. A similar concern about raters applies to measures that rely on observations.

There are many ways to estimate the reliability of scores on a measure, depending on the sources of measurement error that are taken into consideration. Two common indicators of score reliability for school- or classroom-based surveys are (1) test-retest reliability (correlations between scores obtained over two administrations of an assessment) and (2) measures of internal consistency (e.g., the extent to which a student responds similarly across items and typically takes the form of coefficient Alpha). For measures that rely on observations or ratings of open-ended responses, levels of agreement among raters are also used to estimate reliability. These indicators of reliability generally take a value between

0 and 1, with values closer to 1 indicating higher reliability and lower measurement error. An assessment that produces scores with low reliability is typically not useful for decisionmaking because those scores lack the necessary precision to provide a trustworthy indicator.

Reliability: Special Considerations for Climate Assessment

Some reliability considerations are especially relevant to climate assessments. First, the *unit of analysis* for climate surveys is typically not the individual students, staff, or community members who respond to the surveys—the unit of analysis is typically the school or classroom itself. As described previously, in most school-based applications, climate is understood to be a characteristic of the classroom or the school (Marsh, Lüdtke, et al., 2012; Rowan, Raudenbush, and Kang, 1991; Sirotnik, 1980). In the case of climate surveys, it is important not only to understand the precision and reliability of individual scores, but also to understand and investigate the precision and reliability of the class or school aggregates (Kane and Brennan, 1977). In effect, when responding to surveys, teachers, students, or other informants function as judges who rate classroom or school climate (Marsh, Lüdtke, et al., 2012; Bliese, 2000; Chan, 1998; Ma and Willms, 2004; Raudenbush and Jean, 2014). Estimating the reliability of these aggregates involves sources of error that are different from those that might emerge at the level of the individual student (AERA, APA, and NCME, 2014). In this way, it is important to consider at least two sources of measurement error when considering reliability: error from the sampling of persons and error in the sampling of survey items. Importantly, and perhaps not intuitively, reliability is not additive, and just because individual scores are reliable and precise does not mean that the aggregate of those scores (at the classroom or school levels) will also be reliable and precise (Brennan, 1995).²

Observation-based measures have their own complications with regard to estimating reliability. Researchers have suggested that at least three sources of measurement error need to be considered: error from the sampling of lessons (e.g., a teacher

Observation-based measures have their own complications with regard to estimating reliability.

is observed for only a few lessons out of the school year), error from the sampling of raters (e.g., a subset of all possible raters conduct the observations), and error from the sampling of situations (the specific information in the classroom or school that an observer uses to make judgments about climate) (Jaeger, 1993). There is a rich body of evidence showing that even well-trained observers can differ from one another in the severity and leniency of their ratings (Bell et al., 2012; Hill, Charalambous, and Kraft, 2012). Also, there is pervasive evidence that observation-based scores can vary greatly from lesson to lesson (e.g., Cohen and Goldhaber, 2016). However, because student-teacher relationships—an important factor in classroom climate—are based on interactions that are dynamic and that change over time, it is less clear whether variation in observation scores from lesson to lesson should be treated as measurement error or whether they reflect real changes in the quality of relationships (Meyer, Cash, and Mashburn, 2011).

Hidden Sources of Measurement Error

Any source of measurement error that is not explicitly incorporated into reliability analysis is considered a *hidden* source. The most common hidden source of measurement error in school climate surveys is that of occasion. Often, climate surveys are administered only once during the school year. However, perceptions of climate might change from day to day or over the course of the year. If the occasion of survey administration were responsible for measurement error, that error would not be quantifiable with only a single survey administration, and it would not be

mitigated by increasing the number of respondents (Webb and Shavelson, 2005). In school or classroom observations, there are also common hidden sources of measurement error, such as day of the week, time of day, type of classroom activity (e.g., are computer or science labs included in sampling?), grade level, and subject (Bell et al., 2012).

Validity

Although validity is often discussed as if it were an immutable trait of an instrument, validity is actually a feature of how people interpret scores and use them. The term *validity* refers to the degree to which evidence and theory support the interpretation of scores for proposed uses of assessments (AERA, APA, and NCME, 2014). For this reason, validity evidence is required to support or justify each and every use of an assessment (e.g., measuring growth or change, appraising climate disparities across student subgroups, classifying schools, assessing school or teacher accountability). The process of validation entails accumulating evidence supporting proposed score interpretations for specific uses (Kane, 2006). The Standards for Educational and Psychological Testing (AERA, APA, and NCME, 2014) specify four broad sources of validity evidence: (1) evidence based on content, (2) evidence based on response processes, (3) evidence based on internal structure, and (4) evidence based on relations to other variables. Here, we discuss each of these sources in the specific context of measuring school and classroom climate.

1. **Evidence based on content.** *Content* refers to the wording and format of survey items, or the scoring rules for observation domains. Validity evidence based on content can be any information that suggests that the content is appropriate to support the desired interpretations and uses of the measure, such as expert assessment of the constructs being measured in a survey, or expert review of the dimensions and scoring levels for an observation rubric. For example, experts on school climate might be asked to judge the items that are part of a teacher-student relationships scale. The experts would indicate the extent to which

each item aligns with their conceptualization of high-quality teacher-student relationships, point out items that appear to measure something different, and identify important aspects of teacher-student relationships that are not addressed by the proposed items. This expert review could help survey designers decide which items to retain and which to drop (Gehlbach and Brinkworth, 2011).

2. **Evidence based on response processes.**

Investigating the response processes of survey respondents helps to build evidence that items are answered by applying appropriate criteria in rating climate. It is especially important that different subgroups of survey respondents interpret survey items the same way and have similar understandings of how to apply survey scales. This information can be collected, for example, through interviews with respondents as they complete the items. One issue that arises with climate surveys is that relevant subgroups might differ systematically in their ratings of school and classroom climate because of external factors. Individual students could also vary in terms of their standards of comparison (Heine et al., 2002) or the internal scales they use to calibrate their climate perceptions (Guion, 1973). Research has shown persistent differences in school climate perception based on race, gender, and prior academic achievement. Although these systematic differences might reflect meaningful disparities in school-based experiences, these differences also might reflect subjective variations in how individuals within a (sub)population use survey scales. For example, students at higher grade levels might view instructional quality differently from students at lower grade levels because older students have spent more time in schools and have different standards of comparison. This means that teachers who work with older students might receive lower scores than teachers who work with younger students, even if their teaching practices are the same (e.g., West et al., 2017; Heine et al., 2002). In regard to observation protocols, evidence

based on response processes might be gathered by appraising the extent to which raters and observers accurately and consistently apply scoring rules (Bell et al., 2012). Such evidence can be gathered by evaluating raters' agreement with master ratings provided by experts (accuracy) and their agreement with one another (consistency). Gathering such evidence on an ongoing basis can help to mitigate rater drift (Casabianca, Lockwood, and McCaffrey, 2015) and to minimize the implicit biases that raters bring to the observation process.

3. Evidence based on internal structure.

This refers to evidence that items, scales, or domains within the assessment relate to each other in theoretically anticipated ways. For example, in an assessment that claims to measure school safety, teacher-student relationships, and institutional resources, we would expect to see higher correlations among items measuring the same domain (e.g., two items measuring school safety) than among items measuring different domains (e.g., one item measuring school safety and one item measuring teacher-student relationships). Evidence based on internal structure is often collected using factor analysis, which is a statistical tool for exploring relationships among indicators of complex concepts. In the context of survey instruments, factor analysis collapses items into smaller sets of interpretable underlying scales or domains based on the extent to which items covary. It is important to remember that standard factor analysis techniques assume that individual observations are independent. However, this is not the case with climate surveys, in which survey respondents are organized into classrooms or schools and have shared experiences. Under these conditions, conventional factor analysis is likely to produce faulty evidence about internal structure, leading to spurious inferences about the relationships among items, scales, and domains (Julian, 2001; Reise et al., 2005; Schweig, 2014). To ensure that evidence is sound and trustworthy, evidence based on

Although validity is often discussed as if it were an immutable trait of an instrument, validity is actually a feature of how people interpret scores and use them.

internal structure should account for the fact that respondents are organized into classrooms or schools, for example, by using multilevel factor analysis or other techniques that take into account the nested structure of the data (McDonald and Goldstein, 1989; Muthén, 1994). Similar issues arise when standard factor analysis techniques are applied to data from observation protocols. Although it is common to use standard factor analysis techniques to find evidence for internal structure, as was described earlier, observation scores contain multiple sources of error, and not accounting for all the sources of error (particularly rater error) can yield spurious inferences about the relationships among domains on an observation protocol (McCaffrey et al., 2015).

4. Evidence based on relations to other variables. This refers to evidence that survey scales or domains relate to other, conceptually similar scales or domains in theoretically anticipated ways. For example, if a classroom climate survey measures the extent to which a teacher's instruction is challenging, survey-based variables should show positive correlations, with measure of instructional rigor captured using other instruments (e.g., classroom observation protocols). If positive school climate is thought to be associated with improved student achievement, then

survey-derived climate measures should be positively associated with achievement measures (e.g., standardized test scores). The unit of analysis is important to consider when exploring relationships among school and classroom climate–derived variables and other variables. In particular, it is important—and, often, more appropriate—to represent these relationships using multilevel models that can incorporate school or classroom means and control appropriately for the sampling error involved in aggregating individual responses to the school or classroom level (e.g., Marsh, Lüdtke, et al., 2012; Lüdtke, Robitzsch, et al., 2009).

Conclusion

This report is intended to serve as a resource for educators and education policymakers who are interested in monitoring school and classroom climate. Making sense of the wide variety of available measures and of validity and reliability evidence—which is often incomplete or lacking altogether—can be a daunting task, and we encourage educators to draw on such resources as technical assistance centers or local universities that can provide measurement expertise as needed. In addition to providing broad guidance regarding assessment of climate and SEL (Holahan and Batey, 2019; Taylor et al., 2018), we have assembled a list of resources and databases that focus explicitly on climate measures.

Notes

¹This report does not address measures of SEL skills or competencies in detail. Searchable databases of SEL assessments are available from the RAND Corporation (undated) and the Assessment Work Group (Measuring SEL: Using Data to Inspire Practice, undated).

²In addition to conventionally reported reliability indicators like coefficient Alpha, measures like an intraclass correlation coefficient are sometimes reported (Kane and Brennan, 1977; Shrout and Fleiss, 1979) to account for the sampling of persons and to convey some information about the extent to which school or classroom scores are reliable.

List of School Climate Survey Instruments and Resources

In this section, we present resources related to school climate, featuring compendiums of measures and resources for educators and schools. For each entry, we indicate whether validity and reliability evidence is available. However, readers should keep in mind that the evidence supporting specific inferences and uses must be aligned with those inferences and uses and that merely having some validity and reliability statistics might not be sufficient to support use of a particular instrument. Additional guidance regarding interpretation of validity and reliability evidence is available in Taylor et al., 2018.

- “Are You Ready to Assess Social and Emotional Learning and Development?” (American Institutes for Research [AIR])
 - Online tool kit providing guidance for educators and policymakers on establishing the conditions for productive and successful SEL and school climate measurement; includes a list of SEL competency assessments and school climate; no information about validity or reliability provided (AIR, 2019)
- National Center on Safe Supportive Learning Environments “School Climate Survey Compendium” (AIR)
 - Online compendium of survey-based measures of school climate; all surveys have documented evidence of validity and reliability (National Center on Safe Supportive Learning Environments, undated)
- *Measuring School Climate for Gauging Principal Performance* (AIR)
 - Policy report that reviews publicly available measures of school and classroom climate, including information about their technical quality; tool reports available validity and reliability evidence for all documented surveys (Clifford et al., 2012)
- *Making ESSA’s Equity Promise Real: State Strategies to Close the Opportunity Gap* (Learning Policy Institute)
 - Research brief about school climate, ESSA, and state efforts related to school climate (Kostyo, Cardichon, and Darling-Hammond, 2018)
- *Measuring School Climate and Social and Emotional Learning and Development: A Navigation Guide for States and Districts* (CCSSO)
 - Research brief on measuring school climate; includes a list of school climate surveys and resources; no information about validity or reliability is provided (Holahan and Batey, 2019)
- *Lessons from the Field: The Role of Student Surveys in Teacher Evaluation and Development* (Bellwether Education Partners)
 - Describes the experiences of states and local education agencies that have adopted student surveys in teacher feedback and evaluation systems (Schulz, Sud, and Crowe, 2014)
- *Uncommon Measures: Student Surveys and Their Use in Measuring Teaching Effectiveness* (AIR)
 - Research brief providing guidance on the use of student surveys; includes summaries of vendor-developed instruments; reports available validity and reliability evidence for all documented surveys (English et al., 2015)
- *Teaching the Whole Child: Instructional Practices That Support Social-Emotional Learning in Three Teacher Evaluation Frameworks* (AIR)
 - Research brief highlighting how three widely used professional teaching frameworks capture aspects of instruction that promote positive SEL and classroom climate (Yoder, 2014)
- *Classroom Observations and the MET Project* (Bill & Melinda Gates Foundation)
 - Summary highlighting the classroom observation protocols used in the Measures of Effective Teaching Project; tool reports available validity and reliability evidence for all documented surveys (K–12 Education Team, 2010)

References

- AERA, APA, and NCME—See American Educational Research Association, American Psychological Association, and National Council of Measurement in Education.
- AIR—See American Institutes for Research.
- Allensworth, Elaine M., Camille A. Farrington, Molly F. Gordon, David W. Johnson, Kylie Klein, Bronwyn McDaniel, and Jenny Nagaoka, *Supporting Social, Emotional, and Academic Development: Research Implications for Educators*, Chicago, Ill.: University of Chicago Consortium on School Research, 2018. As of September 4, 2019:
<https://consortium.uchicago.edu/sites/default/files/2019-01/Supporting%20Social%20Emotional-Oct2018-Consortium.pdf>
- Allensworth, Elaine M., and Holly Hart, *How Do Principals Influence Student Achievement?* Chicago, Ill.: University of Chicago Consortium on School Research, 2018.
- American Educational Research Association, American Psychological Association, and National Council of Measurement in Education, *Standards for Educational and Psychological Testing*, 2014 ed., Washington, D.C.: American Educational Research Association, 2014.
- American Institutes for Research, “Are You Ready to Assess Social and Emotional Learning and Development?” online tool kit, October 1, 2019. As of November 2, 2019:
<https://www.air.org/resource/are-you-ready-assess-social-and-emotional-learning-and-development-second-edition>
- Artino, Anthony R., Jr., and Hunter Gehlbach, “AM Last Page: Avoiding Four Visual-Design Pitfalls in Survey Development,” *Academic Medicine*, Vol. 87, No. 10, 2012, p. 1452.
- Aspen Institute, *From a Nation at Risk to a Nation at Hope: Recommendations from the National Commission on Social, Emotional, and Academic Development*, Washington, D.C., 2019. As of January 23, 2019:
http://nationathope.org/wp-content/uploads/2018_aspen_final-report_full_webversion.pdf
- Babad, Elisha, “Teachers’ Differential Behavior,” *Educational Psychology Review*, Vol. 5, No. 4, 1993, pp. 347–376.
- Baker, Jean A., Teresa P. Clark, Kimberly S. Maier, and Steve Viger, “The Differential Influence of Instructional Context on the Academic Engagement of Students with Behavior Problems,” *Teaching and Teacher Education*, Vol. 24, No. 7, 2008, pp. 1876–1883.
- Balch, Ryan Thomas, *The Validation of a Student Survey on Teacher Practice*, dissertation, Nashville, Tenn.: Vanderbilt University, 2012.
- Bankston, Carl L., III, and Min Zhou, “Being Well vs. Doing Well: Self-Esteem and School Performance Among Immigrant and Nonimmigrant Racial and Ethnic Groups,” *International Migration Review*, Vol. 36, No. 2, 2002, pp. 389–415.
- Bell, Courtney A., Drew H. Gitomer, Daniel F. McCaffrey, Bridget K. Hamre, Robert C. Pianta, and Yi Qi, “An Argument Approach to Observation Protocol Validity,” *Educational Assessment*, Vol. 17, Nos. 2–3, 2012, pp. 62–87.
- Blad, Evie, “New Online Tool Expands Access to School Climate Measures,” *Education Week*, April 19, 2016. As of September 4, 2019:
<https://www.edweek.org/ew/articles/2016/04/20/new-online-tool-expands-access-to-school.html>
- Bliese, Paul D., “Within-Group Agreement, Non-Independence, and Reliability: Implications for Data Aggregation and Analysis,” in K. J. Klein and S. W. Kozlowski, eds., *Multilevel Theory, Research, and Methods in Organizations*, San Francisco, Calif.: Jossey-Bass, 2000.
- Brennan, Robert L., “The Conventional Wisdom About Group Mean Scores,” *Journal of Educational Measurement*, Vol. 32, No. 4, 1995, pp. 385–396.
- Brophy, Jere E., and Thomas L. Good, *Teacher-Student Relationships: Causes and Consequences*, New York: Holt, Rinehart, and Winston, 1974.
- Burstein, Leigh, Lorraine M. McDonnell, Jeannette Van Winkle, Tor Ormseth, Jim Mirocha, and Gretchen Guiton, *Validating National Curriculum Indicators*, Santa Monica, Calif.: RAND Corporation, MR-658-NSF, 1995. As of November 8, 2019:
https://www.rand.org/pubs/monograph_reports/MR658.html
- Calarco, Jessica McCrory, “‘I Need Help!’ Social Class and Children’s Help-Seeking in Elementary School,” *American Sociological Review*, Vol. 76, No. 6, 2011, pp. 862–882.
- Casabianca, Jodi M., J. R. Lockwood, and Daniel F. McCaffrey, “Trends in Classroom Observation Scores,” *Educational and Psychological Measurement*, Vol. 75, No. 2, 2015, pp. 311–337.
- Cash, Anne H., Bridget Hamre, Robert C. Pianta, and Sonya S. Myers, “Rater Calibration When Observational Assessment Occurs at Large Scale: Degree of Calibration and Characteristics of Raters Associated with Calibration,” *Early Childhood Research Quarterly*, Vol. 27, No. 3, 2012, pp. 529–542.
- Center on Great Teachers and Leaders, “Databases on State Teacher and Principal Evaluation Policies,” webpage, undated. As of November 1, 2019:
<http://resource.tqsource.org/stateevaldb/Compare50States.aspx>
- Centers for Disease Control and Prevention, *School Connectedness: Strategies for Increasing Protective Factors Among Youth*, Atlanta, Ga.: U.S. Department of Health and Human Services, 2009. As of September 9, 2019:
<https://www.cdc.gov/healthyyouth/protective/pdf/connectedness.pdf>
- Chan, David, “Functional Relations Among Constructs in the Same Content Domain at Different Levels of Analysis: A Typology of Composition Models,” *Journal of Applied Psychology*, Vol. 83, No. 2, 1998, pp. 234–246.
- Christle, Christine A., Kristine Jolivet, and C. Michael Nelson, “School Characteristics Related to High School Dropout Rates,” *Remedial and Special Education*, Vol. 28, No. 6, 2007, pp. 325–339.
- Clifford, Matthew, Roshni Menon, Tracy Gangi, Christopher Condon, and Katie Hornung, *Measuring School Climate for Gauging Principal Performance: A Review of the Validity and Reliability of Publicly Accessible Measures*, Washington, D.C.: American Institutes for Research, April 2012. As of November 2, 2019:
https://www.air.org/sites/default/files/downloads/report/school-climate2_0.pdf
- Cohen, Jonathan, Libby McCabe, Nicholas M. Michelli, and Terry Pickeral, “School Climate: Research, Policy, Practice, and Teacher Education,” *Teachers College Record*, Vol. 111, No. 1, 2009, pp. 180–213.

- Cohen, Jonathan, Terry Pickeral, and Molly McCloskey, "Assessing School Climate," *The Education Digest*, Vol. 74, No. 8, 2009, p. 45.
- Cohen, Julie, and Dan Goldhaber, "Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations," *Educational Researcher*, Vol. 45, No. 6, 2016, pp. 378–387.
- Creemers, Bert P. M., *The Effective Classroom*, London: Cassell, 1994.
- Creemers, Bert P. M., and Gerry J. Reezigt, "The Role of School and Classroom Climate in Elementary School Learning Environments," in H. Jerome Freiberg, ed., *School Climate: Measuring, Improving and Sustaining Healthy Learning Environments*, Hampshire: Falmer, 1999, pp. 30–47.
- Croninger, Robert G., and Linda Valli, "'Where Is the Action?' Challenges to Studying the Teaching of Reading in Elementary Classrooms," *Educational Researcher*, Vol. 38, No. 2, 2009, pp. 100–108.
- Deal, Terence E., and Kent D. Peterson, *Shaping School Culture*, Hoboken, N.J.: John Wiley & Sons, 2016.
- DePaoli, Jennifer L., Matthew N. Atwell, and John Bridgeland, *Ready to Lead: A National Principal Survey on How Social and Emotional Learning Can Prepare Children and Transform Schools (A Report for CASEL)*, Washington, D.C.: Civic Enterprises and Hart Research Associates, 2017.
- Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian, *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, Hoboken N.J.: John Wiley & Sons, 2014.
- Downer, Jason T., Megan Stuhlman, Jonathan Schweig, José Feliepe Martínez, and Erik Ruzek, "Measuring Effective Teacher-Student Interactions from a Student Perspective: A Multi-Level Analysis," *Journal of Early Adolescence*, Vol. 35, Nos. 5–6, 2015, pp. 722–758.
- Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger, "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions," *Child Development*, Vol. 82, No. 1, 2011, pp. 405–432.
- Dynarski, Mark, Linda Clarke, Brian Cobb, Jeremy Finn, Russell Rumberger, and Jay Smink, *Dropout Prevention: A Practice Guide (NCEE 2008–4025)*, Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2008.
- English, David, Jackie Burniske, Debra Meibaum, and Lisa Lachlan-Haché, *Uncommon Measures: Student Surveys and Their Use in Measuring Teaching Effectiveness*, Washington, D.C.: American Institutes for Research, November 2015. As of November 2, 2019: <https://www.air.org/sites/default/files/Uncommon-Measures-Student-Surveys-Guidance-Nov-2015.pdf>
- Epstein, Joyce Levy, "The Influence of Friends on Achievement and Affective Outcomes," in Joyce Levy Epstein and Nancy L. Karweit, eds., *Friends in School: Patterns of Selection and Influence in Secondary Schools*, New York: Academic Press, 1983, pp. 177–200.
- Fagell, Phyllis L., "Principal Upset by Staff Response to School Climate Surveys," *Career Confidential*, blog post, *Phi Delta Kappan*, June 26, 2018. As of September 4, 2019: <https://www.kappanonline.org/fagell-principal-upset-staff-response-school-climate-surveys/>
- Fast, Lisa A., James L. Lewis, Michael J. Bryant, Kathleen A. Bocian, Richard A. Cardullo, Michael Rettig, and Kimberly A. Hammond, "Does Math Self-Efficacy Mediate the Effect of the Perceived Classroom Environment on Standardized Math Test Performance?" *Journal of Educational Psychology*, Vol. 102, No. 3, 2010, p. 729.
- Fauth, Benjamin Kasper, and Anna-Katharina Praetorius, *Student, Teacher, and Observer Ratings of Teaching Quality: Investigating Reasons for (Dis)Agreement*, symposium presented at the biennial EARLI Conference for Research on Learning and Instruction, Aachen, Germany, 2019.
- Feldlaufer, Harriet, Carol Midgley, and Jacquelynne S. Eccles, "Student, Teacher, and Observer Perceptions of the Classroom Environment Before and After the Transition to Junior High School," *Journal of Early Adolescence*, Vol. 8, No. 2, 1988, pp. 133–156.
- Ferguson, Ron F., *Student Perceptions of Teaching Effectiveness*, Boston, Mass.: The National Center for Teacher Effectiveness and the Achievement Gap Initiative, 2010.
- , "Can Student Surveys Measure Teaching Quality?" *Phi Delta Kappan*, Vol. 94, No. 3, 2012, pp. 24–28.
- Follman, John, "Secondary School Students' Ratings of Teacher Effectiveness," *High School Journal*, Vol. 75, No. 3, 1992, pp. 168–178.
- Fraser, Barry J., "Classroom Environment Instruments: Development, Validity and Applications," *Learning Environments Research*, Vol. 1, No. 1, 1998, pp. 7–34.
- Freiberg, H. Jerome, "Three Creative Ways to Measure School Climate and Next Steps," in H. Jerome Freiberg, ed., *School Climate: Measuring, Improving and Sustaining Healthy Learning Environments*, London: Falmer Press, 1999.
- Freiberg, H. Jerome, and T. A. Stein, "Measuring, Improving and Sustaining Healthy Learning Environments," *School Climate: Measuring, Improving and Sustaining Healthy Learning Environments*, London: Falmer Press, 1999.
- Fresko, Barbara, and Fadia Nasser, "Interpreting Student Ratings: Consultation, Instructional Modification, and Attitude Towards Course Evaluation," *Studies in Educational Evaluation*, Vol. 27, No. 4, 2001, pp. 291–305.
- Fullan, Michael, "Leadership for the 21st Century: Breaking the Bonds of Dependency," *Educational Leadership*, Vol. 55, 1998, pp. 6–11.
- Garza, Ruben, "Latino and White High School Students' Perceptions of Caring Behaviors: Are We Culturally Responsive to Our Students?" *Urban Education*, Vol. 44, No. 3, 2009, pp. 297–321.
- Gehlbach, Hunter, "Seven Survey Sins," *Journal of Early Adolescence*, Vol. 35, Nos. 5–6, 2015, pp. 883–897.
- Gehlbach, Hunter, and Anthony R. Artino, Jr., "The Survey Checklist (Manifesto)," *Academic Medicine*, Vol. 93, No. 3, 2018, pp. 360–366.
- Gehlbach, Hunter, and Maureen E. Brinkworth, "Measure Twice, Cut Down Error: A Process for Enhancing the Validity of Survey Scales," *Review of General Psychology*, Vol. 15, No. 4, 2011, pp. 380–387.

- Gehlbach, Hunter, Maureen E. Brinkworth, Aaron M. King, Laura M. Hsu, Joseph McIntyre, and Todd Rogers, "Creating Birds of Similar Feathers: Leveraging Similarity to Improve Teacher–Student Relationships and Academic Achievement," *Journal of Educational Psychology*, Vol. 108, No. 3, 2016, p. 342.
- Gehlbach, Hunter, and Heather J. Hough, *Measuring Social Emotional Learning Through Student Surveys in the CORE Districts: A Pragmatic Approach to Validity and Reliability*, Stanford, Calif.: Policy Analysis for California Education, 2018.
- Gehlbach, Hunter, Carly D. Robinson, Ilana Fineffer-Rosenbluh, Chris Benschhoff, and Jack Schneider, "Questionnaires as Interventions: Can Taking a Survey Increase Teachers' Openness to Student Feedback Surveys?" *Educational Psychology*, Vol. 38, No. 3, 2018, pp. 350–367.
- Government Accountability Office, *Race to the Top: States Implementing Teacher and Principal Evaluation Systems Despite Challenges*, Washington, D.C., 2013. As of November 8, 2019: <https://www.gao.gov/assets/660/657936.pdf>
- Gray, Roderic, *A Climate of Success*, London: Routledge, 2007.
- Grissom, Jason A., and Susanna Loeb, "Assessing Principals' Assessments: Subjective Evaluations of Teacher Effectiveness in Low- and High-Stakes Environments," *Education Finance and Policy*, Vol. 12, No. 3, 2017, pp. 369–395.
- Gruenert, Steve, "School Culture, School Climate: They Are Not the Same Thing," *Principal*, March–April 2008, pp. 56–59.
- Guion, Robert M., "A Note on Organizational Climate," *Organizational Behavior and Human Performance*, Vol. 9, No. 1, 1973, pp. 120–125.
- Gurwitz, Sharon B., and Kenneth A. Dodge, "Adults' Evaluations of a Child as a Function of Sex of Adult and Sex of Child," *Journal of Personality and Social Psychology*, Vol. 32, No. 5, 1975, p. 822.
- Hamilton, Laura S., Christopher J. Doss, and Elizabeth D. Steiner, *Teacher and Principal Perspectives on Social and Emotional Learning in America's Schools: Findings from the American Educator Panels*, Santa Monica, Calif.: RAND Corporation, RR-2991-BMGF, 2019. As of July 29, 2019: https://www.rand.org/pubs/research_reports/RR2991.html
- Hamre, Bridget K., and Robert C. Pianta, "Early Teacher–Child Relationships and the Trajectory of Children's School Outcomes Through Eighth Grade," *Child Development*, Vol. 72, No. 2, 2001, pp. 625–638.
- Hanson, Thomas L., and Jin-Ok Kim, "Measuring Resilience and Youth Development: The Psychometric Properties of the Healthy Kids Survey," *REL West Issues & Answers Report*, Vol. 34, 2007.
- Heine, Steven J., Darrin R. Lehman, Kaiping Peng, and Joe Greenholtz, "What's Wrong with Cross-Cultural Comparisons of Subjective Likert Scales? The Reference-Group Effect," *Journal of Personality and Social Psychology*, Vol. 82, No. 6, 2002, p. 903.
- Hill, Heather C., Charalambos Y. Charalambous, and Matthew A. Kraft, "When Rater Reliability Is Not Enough: Teacher Observation Systems and a Case for the Generalizability Study," *Educational Researcher*, Vol. 41, No. 2, 2012, pp. 56–64.
- Hirsch, Eric, and Scott Emerick, *Teaching and Learning Conditions Are Critical to the Success of Students and the Retention of Teachers: Final Report on the 2006 Teaching and Learning Conditions Survey to the Clark County School District and Clark County Education Association*, Carrboro, N.C.: Center for Teaching Quality, 2006.
- Ho, Andrew D., and Thomas J. Kane, *The Reliability of Classroom Observations by School Personnel*, Seattle, Wash.: Bill & Melinda Gates Foundation, MET Project Research Paper, 2013.
- Holahan, Cathy, and Brooklyn Batey, *Measuring School Climate and Social and Emotional Learning and Development: A Navigation Guide for States and Districts*, Washington, D.C.: Council of Chief State School Officers, 2019. As of November 2, 2019: https://ccsso.org/sites/default/files/2019-03/CCSSO-EdCounsel%20SE%20and%20School%20Climate%20measurement_0.pdf
- Hough, Heather, Demetra Kalogrides, and Susanna Loeb, *Using Surveys of Students' Social-Emotional Learning and School Climate for Accountability and Continuous Improvement*, Stanford, Calif.: Policy Analysis for California Education, 2017.
- Hough, Heather, Emily Penner, and Joe Witte, "Identity Crisis: Multiple Measures and the Identification of Schools Under ESSA," Stanford, Calif.: Policy Analysis for California Education, Policy Memo 16–3, 2016.
- Hoy, Wayne K., "Organizational Climate and Culture: A Conceptual Analysis of the School Workplace," *Journal of Educational and Psychological Consultation*, Vol. 1, No. 2, 1990, pp. 149–168.
- Jaeger, Richard M., *Live vs. Memorex: Psychometric and Practical Issues in the Collection of Data on Teachers' Performances in the Classroom*, Greensboro, N.C.: Center for Educational Research and Evaluation, 1993.
- Jagers, Robert J., Deborah Rivas-Drake, and Teresa Borowski, *Equity and Social and Emotional Learning: A Cultural Analysis*, Chicago, Ill.: Establishing Practical Social-Emotional Competence Assessments Work Group, Frameworks Briefs Series, November 2018. As of October 17, 2019: <https://measuringsel.casel.org/wp-content/uploads/2018/11/Frameworks-Equity.pdf>
- Julian, Marc W., "The Consequences of Ignoring Multilevel Data Structures in Nonhierarchical Covariance Modeling," *Structural Equation Modeling*, Vol. 8, No. 3, 2001, pp. 325–352.
- K–12 Education Team, *Classroom Observations and the MET Project*, Seattle Wash.: Bill & Melinda Gates Foundation, 2010.
- Kane, Michael T., "Validity," *Educational Measurement*, Vol. 4, 2006, pp. 17–64.
- Kane, Michael T., and Robert L. Brennan, "The Generalizability of Class Means," *Review of Educational Research*, Vol. 47, No. 2, 1977, pp. 267–292.
- Kane, Thomas J., and Douglas O. Staiger, *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*, Seattle, Wash.: MET Project, Bill & Melinda Gates Foundation, 2012.
- Knapp, Michael S., ed., *Teaching for Meaning in High-Poverty Classrooms*, New York: Teachers College Press, 1995.
- Kostyo, Stephen, Jessica Cardichon, and Linda Darling-Hammond, *Making ESSA's Equity Promise Real: State Strategies to Close the Opportunity Gap*, Palo Alto, Calif.: Learning Policy Institute, 2018.
- Kraft, Matthew A., and Allison F. Gilmour, "Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness," *Educational Researcher*, Vol. 46, No. 5, 2017, pp. 234–249.

- Krosnick, Jon A., "Survey Research," *Annual Review of Psychology*, Vol. 50, No. 1, 1999, pp. 537–567.
- Kunter, Mareike, and Jürgen Baumert, "Who Is the Expert? Construct and Criteria Validity of Student and Teacher Ratings of Instruction," *Learning Environments Research*, Vol. 9, No. 3, 2006, pp. 231–251.
- Kyriakides, Leonidas, "Drawing from Teacher Effectiveness Research and Research into Teacher Interpersonal Behaviour to Establish a Teacher Evaluation System: A Study on the Use of Student Ratings to Evaluate Teacher Behaviour," *Journal of Classroom Interaction*, Vol. 40, 2005, pp. 44–66.
- Little, Olivia, Laura Goe, and Courtney Bell, *A Practical Guide to Evaluating Teacher Effectiveness*, Washington, D.C.: National Comprehensive Center for Teacher Quality, 2009.
- Louis, Karen Seashore, "Beyond 'Managed Change': Rethinking How Schools Improve," *School Effectiveness and School Improvement*, Vol. 5, No. 1, 1994, pp. 2–24.
- Lüdtke, Oliver, Alexander Robitzsch, Ulrich Trautwein, and Mareike Kunter, "Assessing the Impact of Learning Environments: How to Use Student Ratings of Classroom or School Characteristics in Multilevel Modeling," *Contemporary Educational Psychology*, Vol. 34, No. 2, 2009, pp. 120–131.
- Lüdtke, Oliver, Ulrich Trautwein, Mareike Kunter, and Jürgen Baumert, "Reliability and Agreement of Student Ratings of the Classroom Environment: A Reanalysis of TIMSS Data," *Learning Environments Research*, Vol. 9, No. 3, 2006, pp. 215–230.
- Ma, Xin, and J. Douglas Willms, "School Disciplinary Climate: Characteristics and Effects on Eighth Grade Achievement," *Alberta Journal of Educational Research*, Vol. 50, No. 2, 2004.
- MacNeil, Angus J., Doris L. Prater, and Steve Busch, "The Effects of School Culture and Climate on Student Achievement," *International Journal of Leadership in Education*, Vol. 12, No. 1, 2009, pp. 73–84.
- Marsh, Julie A., Susan Bush-Mecenas, Heather J. Hough, Vicki Park, Taylor Allbright, Michelle Hall, and Holly Glover, *At the Forefront of the New Accountability Era: Early Implementation Findings from the CORE Waiver Districts*, Stanford, Calif.: Policy Analysis for California Education, 2016.
- Marsh, Herbert W., Oliver Lüdtke, Benjamin Nagengast, Ulrich Trautwein, Alexandre J. S. Morin, Adel S. Abduljabbar, and Olaf Köller, "Classroom Climate and Contextual Effects: Conceptual and Methodological Issues in the Evaluation of Group-Level Effects," *Educational Psychologist*, Vol. 47, No. 2, 2012, pp. 106–124.
- Mayer, Daniel P., "Measuring Instructional Practice: Can Policymakers Trust Survey Data?" *Educational Evaluation and Policy Analysis*, Vol. 21, No. 1, 1999, pp. 29–45.
- McCaffrey, Daniel F., Kun Yuan, Terrance D. Savitsky, J. R. Lockwood, and Maria O. Edelen, "Uncovering Multivariate Structure in Classroom Observations in the Presence of Rater Errors," *Educational Measurement: Issues and Practice*, Vol. 34, No. 2, 2015, pp. 34–46.
- McDonald, Roderick P., and Harvey Goldstein, "Balanced Versus Unbalanced Designs for Linear Structural Relations in Two-Level Data," *British Journal of Mathematical and Statistical Psychology*, Vol. 42, No. 2, 1989, pp. 215–232.
- Measuring SEL: Using Data to Inspire Practice, *Assessment Guide* registration form, undated. As of November 8, 2019: <http://measuringsel.casel.org/access-assessment-guide/>
- Meyer, J. Patrick, Anne H. Cash, and Andrew Mashburn, "Occasions and the Reliability of Classroom Observations: Alternative Conceptualizations and Methods of Analysis," *Educational Assessment*, Vol. 16, No. 4, 2011, pp. 227–243.
- Mihaly, Kata, and Daniel McCaffrey, "Grade-Level Variation in Observational Measures of Teacher Effectiveness," in Thomas Kane, Kerri A. Kerr, and Robert C. Pianta, eds., *Designing Teacher Evaluation Systems*, New York: Jossey-Bass, 2014, pp. 7–49.
- Mitra, Dana L., "Student Voice in School Reform: From Listening to Leadership," in Dennis Thiessen and Alison Cook-Sather, eds., *International Handbook of Student Experience in Elementary and Secondary School*, Berlin: Springer, Dordrecht, 2007, pp. 727–744.
- Moss, Howard A., and Sandra J. Jones, "Relations Between Maternal Attitudes and Maternal Behavior as a Function of Social Class," in P. Herbert Leiderman, Steven R. Tulkin, and Anne Rosenfeld, eds., *Culture and Infancy: Variations in the Human Experience*, New York: Academic Press, 1977, pp. 439–467.
- Muthén, Bengt O., "Multilevel Covariance Structure Analysis," *Sociological Methods & Research*, Vol. 22, No. 3, 1994, pp. 376–398.
- Nathanson, Lori, Meghan McCormick, and James J. Kemple, *Strengthening Assessments of School Climate: Lessons from the NYC School Survey*, New York: Research Alliance for New York City Schools, 2013.
- National Center on Safe Supportive Learning Environments, "School Climate Survey Compendium," webpage, undated. As of November 2, 2019: <https://safesupportivelearning.ed.gov/topic-research/school-climate-measurement/school-climate-survey-compendium>
- Nayfack, Michelle, Vicki Park, Heather Hough, and Larkin Willis, *Building Systems Knowledge for Continuous Improvement: Early Lessons from the CORE Districts*, Stanford, Calif.: Policy Analysis for California Education, 2017. As of November 1, 2019: <https://edpolicyinca.org/publications/building-systems-knowledge-for-continuous-improvement>
- O'Brennan, Lindsey, and Catherine Bradshaw, *Importance of School Climate*, Washington, D.C.: National Education Association, research brief, 2013.
- Ostroff, Cheri, Angelo J. Kinicki, and Melinda M. Tamkins, "Organizational Climate and Culture," in Irving B. Weiner, Neal W. Schmitt, and Scott Highhouse, eds., *Comprehensive Handbook of Psychology*, Vol. 12: *Industrial and Organizational Psychology*, Hoboken, N.J.: Wiley, 2003, pp. 365–402.
- Peterson, Kenneth D., *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*, Thousand Oaks, Calif.: Corwin Press, 2000.
- Phelan, Patricia, Ann Locke Davidson, and Hanh Cao Yu, *Adolescents' Worlds: Negotiating Family, Peers, and School*, New York: Teachers College Press, 1998.
- Pianta, Robert C., and Bridget K. Hamre, "Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity," *Educational Researcher*, Vol. 38, No. 2, 2009, pp. 109–119.
- Policy Analysis for California Education, "CORE-PACE Research Partnership Publications," webpage, undated. As of November 1, 2019: <https://edpolicyinca.org/projects/core-pace-research-partnership/publications>

- Polikoff, Mogan S., "The Stability of Observational and Student Survey Measures of Teaching Effectiveness," *American Journal of Education*, Vol. 121, No. 2, 2015, pp. 183–212.
- Popham, W. James, *Evaluating America's Teachers: Mission Possible?* Thousand Oaks, Calif.: Corwin Press, 2013.
- Praetorius, Anna-Katharina, and Charalambos Y. Charalambos, "Classroom Observation Frameworks for Studying Instructional Quality: Looking Back and Looking Forward," *ZDM*, Vol. 50, No. 3, 2018, pp. 535–553.
- Price, Heather E., "Assessing U.S. Public School Quality: The Advantages of Combining Internal 'Consumer Ratings' with External NCLB Ratings," *Educational Policy*, Vol. 30, No. 3, 2016, pp. 403–433.
- RAND Corporation, "RAND Education Assessment Finder," webpage, undated. As of November 8, 2019: <https://www.rand.org/education-and-labor/projects/assessments.html>
- Raudenbush, Stephen W., and Marshall Jean, "To What Extent Do Student Perceptions of Classroom Quality Predict Teacher Value Added?" in Thomas J. Kane, Kerri A. Kerr, and Robert C. Pianta, eds., *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*, Hoboken, N.J.: Wiley, 2015, pp. 170–202.
- Reinholz, Daniel L., and Niraj Shah, "Equity Analytics: A Methodological Approach for Quantifying Participation Patterns in Mathematics Classroom Discourse," *Journal for Research in Mathematics Education*, Vol. 49, No. 2, 2018, pp. 140–177.
- Reise, Steven P., Joseph Ventura, Keith H. Nuechterlein, and Kevin H. Kim, "An Illustration of Multilevel Factor Analysis," *Journal of Personality Assessment*, Vol. 84, No. 2, 2005, pp. 126–136.
- Repp, Alan C., Gayla S. Nieminen, Ellen Olinger, and Rita Brusca, "Direct Observation: Factors Affecting the Accuracy of Observers," *Exceptional Children*, Vol. 55, No. 1, 1988, pp. 29–36.
- Rimm-Kaufman, Sara E., and Bridget K. Hamre, "The Role of Psychological and Developmental Science in Efforts to Improve Teacher Quality," *Teachers College Record*, Vol. 112, No. 12, 2010, pp. 2988–3023.
- Roberts, Karlene H., Charles L. Hulin, and Denise M. Rousseau, *Developing an Interdisciplinary Science of Organizations*, San Francisco, Calif.: Jossey-Bass, 1978.
- Rogelberg, Steven G., and Alexandra Luong, "Nonresponse to Mailed Surveys: A Review and Guide," *Current Directions in Psychological Science*, Vol. 7, No. 2, 1998, pp. 60–65.
- Rogelberg, Steven G., and Jeffrey M. Stanton, "Introduction: Understanding and Dealing with Organizational Survey Nonresponse," *Organizational Research Methods*, Vol. 10, No. 2, 2007, pp. 195–209.
- Rowan, Brian, and Richard Correnti, "Studying Reading Instruction with Teacher Logs: Lessons from the Study of Instructional Improvement," *Educational Researcher*, Vol. 38, No. 2, 2009, pp. 120–131.
- Rowan, Brian, Stephen W. Raudenbush, and Sang Jin Kang, "Organizational Design in High Schools: A Multilevel Analysis," *American Journal of Education*, Vol. 99, No. 2, 1991, pp. 238–266.
- Rowan, Brian, Stephen G. Schilling, Angeline Spain, Prem Bhandari, Daniel Berger, and John Graves, *Promoting High Quality Teacher Evaluations in Michigan: Lessons from a Pilot of Educator Effectiveness Tools*, Ann Arbor: University of Michigan Institute for Social Research, 2013.
- Sarason, Seymour Bernard, *The Culture of the School and the Problem of Change*, Boston, Mass.: Allyn and Bacon, 1982.
- Schoenfeld, Alan H., "Classroom Observations in Theory and Practice," *ZDM*, Vol. 45, No. 4, 2013, pp. 607–621.
- Schonert-Reichl, Kimberly A., "Social and Emotional Learning and Teachers," *Future of Children*, Vol. 27, No. 1, 2017, pp. 137–155.
- Schulz, Jeff, Gunjan Sud, and Becky Crowe, *Lessons from the Field: The Role of Student Surveys in Teacher Evaluation and Development*, Washington, D.C.: Bellwether Education Partners, 2014. As of November 2, 2019: https://bellwethereducation.org/sites/default/files/Bellwether_StudentSurvey.pdf
- Schweig, Jonathan, "Cross-Level Measurement Invariance in School and Classroom Environment Surveys: Implications for Policy and Practice," *Educational Evaluation and Policy Analysis*, Vol. 36, No. 3, 2014, pp. 259–280.
- Shindler, John, Albert Jones, A. Dee Williams, Clint Taylor, and Hermenia Cardenas, "The School Climate-Student Achievement Connection: If We Want Achievement Gains, We Need to Begin by Improving the Climate," *Journal of School Administration Research and Development*, Vol. 1, No. 1, 2016, pp. 9–16.
- Shrout, Patrick E., and Joseph L. Fleiss, "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychological Bulletin*, Vol. 86, No. 2, 1979, p. 420.
- Sirotnik, Kenneth A., "Psychometric Implications of the Unit-of-Analysis Problem (with Examples from the Measurement of Organizational Climate)," *Journal of Educational Measurement*, Vol. 17, No. 4, 1980, pp. 245–282.
- Smircich, Linda, "Is the Concept of Culture a Paradigm for Understanding Organizations and Ourselves," in Peter J. Frost, Larry Moore, Meryl Reis Louis, Craig C. Lundberg, and Joanne Martin, eds., *Organizational Culture*, Thousand Oaks, Calif.: Sage, 1985.
- Steinberg, Matthew P., and Rachel Garrett, "Classroom Composition and Measured Teacher Performance: What Do Teacher Observation Scores Really Measure?" *Educational Evaluation and Policy Analysis*, Vol. 38, No. 2, 2016, pp. 293–317.
- Stevens, Carla J., and Kathryn S. Sanchez, "Perceptions of Parents and Community Members as a Measure of School Climate," in H. Jerome Freiberg, ed., *School Climate: Measuring, Improving, and Sustaining Healthy Learning Environments*, London: Routledge, 1999.
- Stevens, Dannelle D., and Antonia J. Levi, *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*, Sterling, Va.: Stylus Publishing, 2005.
- Stodolsky, Susan S., "Classroom Observation," in Jason Millman and Linda Darling-Hammond, eds., *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers*, Thousand Oaks, Calif.: Corwin, 1990, pp. 175–190.
- Stolp, Stephen, and Stuart C. Smith, *Transforming School Culture: Stories, Symbols, Values & the Leader's Role*, Ann Arbor, Mich.: ERIC Clearinghouse on Educational Management, 1995.

Stringfield, Sam, "The Phoenix Rises from Its Ashes . . . Doesn't It?" in H. Jerome Freiberg, ed., *School Climate: Measuring, Improving, and Sustaining Healthy Learning Environments*, London: Falmer Press, 1999.

Taylor, Jeremy J., Katie Buckley, Laura S. Hamilton, Brian M. Stecher, Lindsay Read, and Jonathan Schweig, *Choosing and Using SEL Competency Assessments: What Schools and Districts Need to Know*, Chicago, Ill.: Collaborative for Academic, Social, and Emotional Learning, 2018. As of January 23, 2019: http://measuringSEL.caseli.org/pdf/Choosing-and-Using-SEL-Competency-Assessments_What-Schools-and-Districts-Need-to-Know.pdf

Thapa, Amrit, Jonathan Cohen, Shawn Guffey, and Ann Higgins-D'Alessandro, "A Review of School Climate Research," *Review of Educational Research*, Vol. 83, No. 3, September 2013, pp. 365–366. As of January 23, 2019: <http://rer.sagepub.com/content/83/3/357.full.pdf+html>

TNTP, *Greenhouse Schools: How Schools Can Build Cultures Where Teachers and Students Thrive*, Brooklyn, N.Y., 2012.

Toch, Thomas, and Raegen Miller, "CORE Lessons: Measuring the Social and Emotional Sides of Student Success," *FutureEd*, February 2, 2019. As of September 4, 2019: <https://www.future-ed.org/core-lessons-measuring-the-social-and-emotional-dimensions-of-student-success/>

Tripod, homepage, undated. As of November 1, 2019: <https://tripoded.com>

U.S. Department of Education, "Topic Areas: The EDSCS School Climate Model," webpage, undated. As of November 1, 2019: <https://safesupportivelearning.ed.gov/edscls/measures#Topic%20Areas%20the%20EDSCS%20Measures>

Van Houtte, Mieke, "Climate or Culture? A Plea for Conceptual Clarity in School Effectiveness Research," *School Effectiveness and School Improvement*, Vol. 16, No. 1, 2005, pp. 71–89.

Voight, Adam, Thomas Hanson, Meagan O'Malley, and Latifah Adekanye, "The Racial School Climate Gap: Within-School Disparities in Students' Experiences of Safety, Support, and Connectedness," *American Journal of Community Psychology*, Vol. 56, Nos. 3–4, 2015, pp. 252–267.

Wang, Margaret C., Geneva D. Haertel, and Herbert J. Walberg, "Toward a Knowledge Base for School Learning," *Review of Educational Research*, Vol. 63, No. 3, 1993, pp. 249–294.

Wang, Ming-Te, and Jessica L. Degol, "School Climate: A Review of the Construct, Measurement, and Impact on Student Outcomes," *Educational Psychology Review*, Vol. 28, No. 2, 2016, pp. 315–352.

Webb, Noreen M., and Richard J. Shavelson, "Generalizability Theory: Overview," *Encyclopedia of Statistics in Behavioral Science*, Vol. 2, 2005, pp. 717–719.

West, Martin R., Katie Buckley, Sara Bartolino Krachman, and Noah Bookman, "Development and Implementation of Student Social-Emotional Surveys in the Core Districts," *Journal of Applied Developmental Psychology*, Vol. 55, 2017, pp. 119–129.

Whitehurst, Grover J., Matthew M. Chingos, and Katharine M. Lindquist, *Evaluating Teachers with Classroom Observations*, Washington, D.C.: Brown Center on Education Policy, Brookings Institution, 2014.

Worrell, Frank C., and Laura D. Kuterbach, "The Use of Student Ratings of Teacher Behaviors with Academically Talented High School Students," *Journal of Secondary Gifted Education*, Vol. 12, No. 4, 2001, pp. 236–247.

Yoder, Nicholas, *Teaching the Whole Child: Instructional Practices That Support Social-Emotional Learning in Three Teacher Evaluation Frameworks*, Washington, D.C.: Center on Great Teachers and Leaders, American Institutes for Research, January 2014. As of January 23, 2019: <https://gtlcenter.org/sites/default/files/TeachingtheWholeChild.pdf>

About This Report

High-quality measures of school and classroom climate can help educators monitor and improve students' learning environments, but educators frequently lack access to information about what measures are available or about how to use data from these measures in ways that will support improvement efforts. This report is intended to serve as a resource for educators and education policymakers who are interested in monitoring school and classroom climate. It provides considerations for using climate measures and includes a list of online sources where educators and policymakers can obtain access to measures. This report serves as a companion document to the RAND Education Assessment Finder (www.rand.org/education-and-labor/projects/assessments), which provides a searchable database of student assessments.

We are grateful to the Einhorn Family Charitable Trust, the William and Flora Hewlett Foundation, the Overdeck Family Foundation, and The Wallace Foundation for supporting this work. We also thank Alice Huguet, Hunter Gehlbach, Samantha Neiman, Deborah Moroney, and Fatih Unlu for their helpful and constructive reviews of earlier drafts of this report, and Arwen Bicknell and Monette Velasco, who provided editing and design assistance.

This report is intended to serve as a resource for educators and education policymakers who are interested in monitoring school and classroom climate.

This publication was commissioned as part of the work of the Funders Collaborative for Innovative Measurement, which was formed in 2015 to advance measurement in interpersonal and intrapersonal skills. In 2019, it was absorbed into Grantmakers for Thriving Youth. For more RAND publications from this effort, visit www.rand.org/education-and-labor/projects/assessments.

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking.

More information about RAND can be found at www.rand.org. Questions about this report should be directed to jrschweig@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.



The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

For more information on this publication, visit www.rand.org/t/RR4259.

© Copyright 2019 RAND Corporation

www.rand.org