Measuring School Climate for Gauging Principal Performance:
A Review of the Validity and Reliability of Publicly Accessible Measures

March 2012

Matt Clifford
Roshni Menon
Tracy Gangi
Chris Condon
Katie Hornung

Copyright © 2012 American Institutes for Research. All rights reserved.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>School Climate and Principal Effectiveness</td>
<td>2</td>
</tr>
<tr>
<td>Procedure</td>
<td>5</td>
</tr>
<tr>
<td>School Climate Surveys Reviewed</td>
<td>7</td>
</tr>
<tr>
<td>Alliance for the Study of School Climate–School Climate Assessment Inventory</td>
<td>7</td>
</tr>
<tr>
<td>Brief California School Climate Survey</td>
<td>7</td>
</tr>
<tr>
<td>Comprehensive Assessment of Leadership for Learning</td>
<td>7</td>
</tr>
<tr>
<td>Comprehensive School Climate Inventory</td>
<td>8</td>
</tr>
<tr>
<td>Creating a Great Place to Learn Survey</td>
<td>8</td>
</tr>
<tr>
<td>Culture of Excellence and Ethics Assessment</td>
<td>8</td>
</tr>
<tr>
<td>The Essential Supports for School Improvement</td>
<td>8</td>
</tr>
<tr>
<td>Inventory of School Climate—Teacher</td>
<td>9</td>
</tr>
<tr>
<td>The Teacher Version of My Class Inventory—Short Form</td>
<td>9</td>
</tr>
<tr>
<td>School Climate Inventory—Revised</td>
<td>9</td>
</tr>
<tr>
<td>Teaching Empowering Leading and Learning Survey</td>
<td>10</td>
</tr>
<tr>
<td>Findings</td>
<td>11</td>
</tr>
<tr>
<td>Conclusion</td>
<td>12</td>
</tr>
<tr>
<td>References</td>
<td>18</td>
</tr>
</tbody>
</table>
Introduction

Many states and school districts are working to improve principal performance evaluations as a means of ensuring that effective principals are leading schools. Federal incentive programs (e.g., Race to the Top, the Teacher Incentive Fund, and School Improvement Grants) and state policies support consistent and systematic measurement of principal effectiveness so that school districts can clearly determine which principals are most and least effective and provide appropriate feedback for improvement. Although professional standards are in place to clearly articulate what principals should know and do, states and school districts are often challenged to determine how to measure principal performance in ways that are useful, valid, and reliable.

Designing principal performance evaluation is challenging for several reasons, two of which are given here. First, the literature provide little guidance on effective principal performance evaluation models. Few research or evaluation studies have been conducted to test the effectiveness of one evaluation system over another (Clifford and Ross, 2011; Sanders and Kearney, 2011). Second, reliance on current practice is also problematic. Goldring, Cravens, Porter, Murphy, and Elliot (2007) found that district performance evaluation practices are inconsistent and idiosyncratic and provide little meaningful feedback to improve leadership practice. This means that performance evaluation designers have few guideposts to inform new designs.

One guidepost offered by research suggests that principals influence teaching and learning by creating a safe and supportive school climate. Some designers of improved school principal evaluation systems are including school climate surveys as one of many measures of principal performance in a single evaluation system. School climate data are important sources of feedback because principals often have control over school-level conditions, although they have less direct control over classroom instruction or teaching quality (Hallinger & Heck, 1998). High-quality principal performance evaluations are closely aligned to educators’ daily work and immediate spheres of influence (Joint Committee on Standards for Educational Evaluation, 2010). Such evaluation data offer educators opportunities to reflect on and improve their practices.

This policy brief provides principal evaluation system designers information about the technical soundness and cost (i.e., time requirements) of publicly available school climate surveys. We focus on the technical soundness of school climate surveys because we believe that using validated and reliable surveys as an outcomes measure can contribute to an evaluation’s fairness, accuracy, and utility for a state or a school district. However, none of the climate surveys that we reviewed were expressly validated for principal evaluation purposes. We advise states and school districts to carefully study principal evaluation systems that are performing well and then select climate surveys that are useful measures of performance.

In addition, policymakers tell us that they need technical soundness and cost information to initially screen possible measures for inclusion in principal evaluation systems. Designers can

---

1 For available measures, see the Guide to Evaluation Products at http://resource.tqsource.org/GEP for an unreviewed list of principal evaluation products.
use the information presented in this brief to identify technically sound school climate surveys and then critically review those surveys to determine how well they fit into principal evaluation system designs.

This brief begins by overviewsing school climate surveys and their potential uses for principal evaluation. Next it outlines our procedure for reviewing school climate surveys, which is followed by brief synopses of each survey that meets the minimum criteria for inclusion in the review. The brief ends with a discussion of the surveys reviewed.

**School Climate and Principal Effectiveness**

Policymakers and educators might take the view that principals, as school leaders, are ultimately responsible for all that occurs in a school building and all aspects of organizational performance. Such a perspective raises the possibility that principals will be evaluated on things that they do not (or cannot) control. Effective performance evaluations focus on aspects of school life and learning that principals can reasonably affect. For example, principals in some school districts have no budget allocation authority; therefore, it makes little sense to evaluate their performance as budget developers or managers. Performance feedback resulting from evaluations closely tied to work practices is more useful for changing practice than those that are not well aligned (DeNisi & Kruger, 2000; Joint Committee on Standards for Educational Evaluation, 2010).

Although the work responsibilities of principals vary, measures of school conditions offer principals and their evaluators useful feedback on performance because principals tend to have a direct influence on school conditions. Principals tend to have authority in controlling school-level conditions, such as school climate, and principals influence student learning by creating conditions within a school for better teaching and learning to occur. Studies of principal influence on student achievement note that their influence is indirect, meaning that principals affect student learning through the work of others. Principals, through their leadership and management practices, can

---

**Box 1. Climate, Culture and Context: What’s the Difference?**

The terms *climate*, *culture*, and *context* are frequently used interchangeably in education, but some argue that differences exist between these constructs (Deal & Peterson, 1999). Each term has different meanings, and no set list of variables is assigned to each term.

For the purposes of this brief, we define *climate* as the quality and the characteristics of school life, which includes the availability of supports for teaching and learning. It includes goals, values, interpersonal relationships, formal organizational structures, and organizational practices.

*Culture* refers to shared beliefs, customs, and behaviors. Culture can be measured, but school culture measures are not included in this brief. Culture represents people’s experiences with ceremonies, beliefs, attitudes, history, ideology, language, practices, rituals, traditions, and values.

*Context* is the conditions surrounding schools, which interact with the culture and the climate in a school. School context can be measured, but such measures are not included in this brief.
determine what human, financial, material, and social resources are brought to bear on schools, and how those resources are allocated (Hallinger & Heck, 1998; Leithwood, Louis, Anderson, & Wahlstrom; 2004). These functions are reflected in national professional standards for principals (e.g., Council of Chief State School Officers, 2008; National Association of Elementary School Principals, 2009).

For these and other reasons, states and school districts have turned to school surveys—specifically school climate surveys—as measures of principal performance (see Box 1 for a definition of school climate). As research indicates, school climate is associated with robust and encouraging outcomes, such as better staff morale (Bryk & Driscoll, 1988) and greater student academic achievement (Shindler, Jones, Williams, Taylor, & Cardenas, 2009). Conversely, school climate research has indicated that a poor school climate is associated with higher absenteeism (Reid, 1983), suspension rates (Wu, Pink, Crain, & Moles, 1982), and school dropout rates (Anderson, 1982). Mowday, Porter, and Steers (1982) and Wynn, Carboni, and Patall (2007) reported that schools with negative school climates had high teacher absenteeism and turnover.

School climate surveys have a long history of use in education and educational research, but only recently have they been used for principal evaluation. For example, researchers have used climate surveys to determine whether school improvement efforts have achieved the desired effects or explain why some schools perform better or worse than others (e.g., a shared mission or vision). Climate surveys meet these purposes by asking teachers, staff, and others to make judgments about a school. A climate survey, for example, might ask teachers about how much they trust their colleagues, how much they believe in the school mission, or how safe they feel in expressing their ideas and opinions.

School climate surveys differ from school audits or school walk-throughs, which are administered by personnel external to a school and also focus on the relative presence of school-level traits. School climate surveys also differ from 360-degree assessments of principal practice because these assessments focus exclusively on gathering multiple perspectives on a principal’s performance at a single point in time. (For a review of leadership assessments, see Condon & Clifford [2010].) Climate surveys more broadly assess the quality and the characteristics of school life, which include the availability of supports for improved teaching and learning.

State-level staff, superintendents, and others seeking to redesign performance evaluation systems may determine that a school climate survey should be used as one measure of principal effectiveness and seek existing surveys or develop one of their own. (For an overview of the principal evaluation design process, see Clifford et al. [2011].) Climate surveys with publicly available testing and validation information can provide policymakers the data they need to make decisions about which climate survey is best for their state or school district. Validated and reliable surveys can contribute to the fairness, the accuracy, and the legal defensibility of performance evaluation systems.

School climate is considered an outcome or a result of principals’ work, such as improved instructional quality, community relationships, or student growth. When used for principal evaluation purposes, school climate surveys can contribute to a summative evaluation of principal performance; yet they are also often used for formative purposes. Principal evaluation
system designers must determine if and how school climate measures should be included in principal evaluation, including what priority such results should be given in a principal’s overall evaluation. Principal evaluation systems also include measures of principal practice, such as observations or a portfolio review, which provide insights on the quality of the work of principals.

**Box 2. Definitions of Reliability and Validity**

The school climate surveys included in this brief have not been validated for the purpose of principal evaluation, but several are currently being used for that purpose. We believe that using validated school climate surveys contributes to the credibility of principal evaluation systems, and policymakers should consider validity and reliability as criteria for selecting measures for the principal evaluation system. Subsequent to the selection of measures, policymakers will need to examine—and possibly adapt—surveys for use as principal evaluation instruments.

For an instrument to be included in this review, technical information on the psychometric soundness (i.e., accepted tested measures used to test for reliability and validity) on the instrument had to be publicly available, either on the Internet or by request. The instrument was determined to be psychometrically sound through a set of a priori criteria developed by the research panel, which will be discussed in the “Procedure” that follows. Reliability and validity testing provides evidence of psychometric rigor when measuring school climate. Psychometric rigor must be considered prior to implementing such instruments in schools and school districts, to ensure that the information gathered is accurate and valid, because of the high stakes in principal performance evaluations.

*Reliability* is the extent to which a measure produces similar results when repeated measurements are made. For example, if an instrument is used to measure school climate, it should consistently produce the same results as long as the school climate and the survey respondents have not changed.

*Validity* is the extent to which an instrument measures what it is intended to measure. This review was concerned with two types of validity: construct validity and content validity. *Construct validity* is an instrument’s ability to identify or measure the variables or the constructs that it proposes to identify or measure. For example, if an instrument intends to measure school safety as one construct of school climate, then multiple items on the survey instrument are needed to measure the degree to which a school is safe. Testing the construct validity of the school safety construct would determine how well the survey items measure school safety.

*Content validity* is the degree to which the content of the items within a survey instrument accurately reflects the various facets of a content domain or a construct. To use the same example as earlier, if school safety is one construct that a survey instrument intends to measure, then items within the instrument need to cover aspects of school safety (e.g., drug use and violence) that are identified in the research literature, by an expert review panel, or a set of widely accepted research-based standards.

The review does not include other forms of validity because we believed these forms of validity are a good starting point for policymakers’ deliberations. Other forms of validity are important, and we encourage policymakers to review all the available literature on the surveys prior to selection.
Procedure

To prepare this brief, we reviewed technical information on publicly available school climate surveys. All the surveys in this review were created and distributed by private companies, higher education institutions, school districts, or states for either no cost or a fee. The survey developers provided technical information on the survey contents, time requirements, and psychometric testing either through promotional literature, peer-reviewed journals, or on request. Some surveys were excluded from this brief because the content was proprietary or the survey developers did not publicly provide technical information with which to judge validity and reliability. Our purpose for the review is to provide principal evaluation system designers information about available school climate surveys. We do not endorse any particular survey.

We conducted a keyword\textsuperscript{2} search of Google Scholar and Google to locate instruments measuring school climate. When the initial search yielded about 1,000 leads to follow, additional keywords of reliability and validity were added, and a content screen was conducted. The content screen narrowed the review to school climate surveys focusing on key aspects of school leadership, rather than a single and constrained focus. For example, we excluded surveys that focused solely on school safety because school leadership involves more than school safety, and we assumed that states and school districts would be disinclined to administer multiple surveys to assess principal performance.

This reduced the number of leads returned to about 125, which included but was not limited to surveys located by Gangi (2010) and AIR’s Safe and Supportive Schools Technical Assistance Center (2011). Both sources were located after the initial review, and both independently analyzed psychometric properties of publicly available surveys through expert review. In addition, a snowball sampling method was used to query experts on principal performance evaluation design at state and district levels (Miles & Huberman, 1994).

The criteria for sampling instruments were as follows:

- Reports on school climate as an intended use of the instrument.
- The instrument and the technical reporting information must be publicly available either on the Internet or by request.
- The instrument was developed within the last 11 years, which avoids the consideration of older instruments that do not reflect the dynamic nature of school climate research.
- The instrument is psychometrically sound (in reliability and validity) according to a priori criteria set by the research panel.
- Either teachers or school administrators complete the instrument.\textsuperscript{3}

Our work is ongoing, and we welcome the opportunity to conduct additional, impartial reviews of other school climate surveys that are being used.

\textsuperscript{2} The keywords used were school climate survey, measuring school environment, and school learning environment.
\textsuperscript{3} Parent and student surveys can also be administered to gauge school climate, but these surveys were excluded from the review because they are less commonly used for principal evaluation.
For the purposes of the review, psychometrically sound means that the instrument must be tested for validity and reliability using accepted testing measures. A minimum overall scale for reliability rating of .75 must be achieved. Also, content validity must have been evidenced by, at minimum, a rigorous literature review or an expert panel review. Finally, construct validity testing must be adequately documented to allow the research panel to judge the relative rigor by which the testing occurred (see Box 2).

Using these criteria, about 25 instruments were initially identified, but only 11 instruments met all the criteria to be included in the final review (see Table 1). Two American Institutes for Research (AIR) researchers conducted separate full reviews of the identified school climate surveys, and two additional AIR reviewers served as objective observers of the review process.

---

4 The research community differs on the benchmark or the minimum scale reliability needed to signify a reliable measure. Some researchers use a benchmark of .70, while others use .80 as a minimum. We set a minimum overall scale reliability of .75 as a compromise. Sometimes, overall reliability indexes are not computed for a scale because some researchers believe that an overall reliability for a scale that includes several subscales measuring different constructs is not warranted. In cases where no overall scale reliability is reported for the measures in this document, we report on the average subscale reliability. We set a minimum average subscale reliability of .60 for inclusion in this table, which is smaller than the benchmark for overall scale reliability, because subscales have fewer items, leading to smaller reliability coefficients. If the overall scale and subscale reliabilities are reported, we considered the overall scale reliability to be more definitive.
School Climate Surveys Reviewed

Alliance for the Study of School Climate–School Climate Assessment Inventory

Shindler, Taylor, Cadenas, and Jones (2003) originally developed the Alliance for the Study of School Climate–School Climate Assessment Inventory (ASSC–SCAI), which was published in 2004 by the Western Alliance for the Study of School Climate (now the Alliance for the Study of School Climate). According to Shindler et al. (2009), SCAI’s purpose is to capture a detailed understanding of each school’s function, health, and performance. It provides surveys for faculty, parents, and students for elementary, middle, and high schools that can be administered either individually or in a group setting. It takes approximately 20 minutes to complete. The measured constructs are physical appearance, faculty relations, student interactions, leadership and decisions, discipline environment, learning and assessment, attitude and culture, and community relations. For more information on ASSC–SCAI, see http://www.calstatela.edu/centers/schoolclimate/assessment/school_survey.html.

Brief California School Climate Survey

You, O’Malley, and Furlong (2012) developed the Brief California School Climate Survey (BCSCS) in response to the data collection requirement within the Safe and Drug Free Schools and Communities Act. BCSCS is adapted from California School Climate Survey and provides schools with data that can be used to promote a healthy learning and working environment. The survey is administered to teachers, administrators, and other school staff, and the responses are completed and submitted electronically. The completion time is not reported; based on the number of items, we estimate it will take about 7–10 minutes. It measures two major constructs: relational supports and organizational supports. For more information, see You et al. (2012) or http://cscs.wested.org/faqs_outside_ca.

Comprehensive Assessment of Leadership for Learning

The Wisconsin Center for Educational Research (WCER) created the Comprehensive Assessment of Learning for Learning (CALL). This survey instrument is the only survey reviewed for this brief that was developed as a formative feedback tool for school leaders. CALL gathers data from principals, school staff, and teachers and is intended for middle schools and high schools. WCER will soon be developing an elementary school version. The completion time for this survey is approximately 45-60 minutes. CALL captures current leadership practices in five domains: focus on learning, monitoring teaching and learning, building nested learning communities, acquiring and allocating resources, and maintaining a safe and effective learning environment. The survey is administered online to administrators and instructional staff and includes an electronic analysis and reporting mechanism. CALL also includes a procedure for ensuring respondent anonymity through the assignment of special codes. For more information, see Halverson, Kelley, and Dikkers (2010) or http://www.callsurvey.org.
Comprehensive School Climate Inventory

The Center for Social and Emotional Education (CSEE; now the National School Climate Center) developed the Comprehensive School Climate Inventory (CSCI) in 2002 to measure the strengths and the needs of schools by surveying students, parents, and school staff. CSCI has versions available for elementary, middle, and high schools, and the reported completion time is 15–20 minutes. The measured constructs fall under four broad categories: safety, teaching and learning, interpersonal relationships, and institutional environment. The school staff version of the survey measures two additional constructs: leadership and professional relationships. For more information on CSCI, see http://www.schoolclimate.org.

Creating a Great Place to Learn Survey

Developed by the Search Institute (2006), the Creating a Great Place to Learn (CGPL) Survey focuses on the psychosocial and learning environment as experienced by students and staff. The student survey measures 11 dimensions, and the staff survey measures 17 dimensions. The dimensions can be organized into the following 3 categories: relationships, organizational attributes, and personal development. The completion time is not provided; based on the number of items, we estimate it will take about 30 minutes to complete the student survey and about 40 minutes to complete the faculty and staff survey. For more information, see Search Institute (2006) or http://www.search-institute.org/survey-services/surveys/creating-great-place-learn.

Culture of Excellence and Ethics Assessment

The Culture of Excellence and Ethics Assessment (CEEA) from the Institute for Excellence and Ethics is a comprehensive battery of school climate survey tools for students, staff, and parents. It focuses on the cultural assets, or protective factors, provided by school and family culture. The completion time is not provided; based on the number of items, we estimate the student survey to take between 35 and 40 minutes, the staff survey to take between 45 and 50 minutes, and the parent survey to take about 25 minutes to complete. The student and faculty-staff surveys include 3 constructs (with additional subconstructs): safe, supportive, and engaging climate; culture of excellence; and ethics. The faculty-staff survey includes a fourth construct for professional community and school-home partnership. For more information, see http://excellenceandethics.com/assess/ceea.php.

The Essential Supports for School Improvement

The Consortium on Chicago School Research developed the Essential Supports for School Improvement to measure a school’s climate against the five evidence-based, essential supports necessary for school improvement and student learning. The Chicago Public Schools used the framework as a guide for school improvement planning and self-assessment efforts and has developed an electronic interface to support administration and data analysis. The five essential supports are leadership, parent-community ties, professional capacity, student-centered learning climate, and curriculum alignment. The survey is taken by teachers, students, and principals. The completion time is not reported; based on the number of items, we estimate it to take approximately 90 minutes to complete the teacher survey. For more information, see Sebring,
Allensworth, Bryk, Easton, and Luppescu (2006) or http://ccsr.uchicago.edu/content/publications.php?pub_id=86.

**Inventory of School Climate—Teacher**

Brand, Felner, Seitsinger, Burns, and Bolton (2008) developed the Inventory of School Climate-Teacher (ISC-T) to collect information on teachers’ views of school climate to understand the effect of school climate on school functioning and school reform efforts. The survey is completed by teachers and measures 6 dimensions: peer sensitivity, disruptiveness, teacher-pupil interactions, achievement orientation, support for cultural pluralism, and safety problems. The completion time is not reported; based on the number of items, we estimate that it will require 15–20 minutes to complete. For more information on ISC-T, see Brand et al. (2008).

**The Teacher Version of My Class Inventory—Short Form**

Sink and Spencer (2007) developed the Teacher Version of My Class Inventory—Short Form as an accountability measure for elementary school counselors to use when evaluating a school’s counseling program. This instrument assesses teachers’ perceptions of the classroom climate as they relate to 5 scales: overall student satisfaction with the learning experience, peer relations, difficulty level of classroom materials, student competitiveness, and school counselor impact on the learning environment. The completion time is not reported; based on the number of items, we estimate it will take approximately 12–15 minutes to complete. For more information, see Sink and Spencer (2007).

**School Climate Inventory-Revised**

The School Climate Inventory-Revised (SCI-R) was originally developed to determine the effect of school reform efforts. Dean Butler and Martha Alberg (Butler & Alberg, 1991) developed SCI-R for the Center for Research in Educational Policy (CREP) at the University of Memphis. It was published in 1989, and revised in 2002. According to the authors, the survey provides formative feedback to school leaders on personnel perceptions of climate and identifies potential interventions specifically for the climate factors that hinder a school’s effectiveness. The instrument surveys faculty and is intended to be administered in a group setting over a 20-minute period. The measured constructs are order, leadership, environment, involvement, instruction, expectations, and collaboration. For additional information on contractual arrangements for SCI-R administration or use, contact CREP at 901-678-2310 or 1-866-670-6147.

**Teaching Empowering Leading and Learning Survey**

The Teaching Empowering Leading and Learning (TELL) Survey was published by the New Teacher Center in 2002; and revised in 2011. The revised survey measures 8 constructs: time, facilities and resources, community support and involvement, managing student conduct, teacher leadership, school leadership, professional development, and instructional practices and support. Each construct contains numerous items; states can add, delete, or revise items to align the survey with their specific context. The survey is administered electronically through a centralized hub administered by the New Teacher Center, which provides survey access, data
displays, and supportive text to assist with date interpretation. The survey is being used for principal evaluation purposes by states and school districts. The completion time is not reported; based on the number of items, we estimate it will take approximately 20 minutes to complete the survey. For more information, see http://www.newteachercenter.org/node/1359.
Discussion

School principals are responsible for creating a school climate that is amenable to teaching and learning improvement. Policymakers are, logically, investigating school climate surveys as a means to evaluate principals. As policymakers consider measurement options, we believe that they should critically review school climate surveys for technical soundness (validity and reliability) and cost. Valid and reliable climate surveys can contribute to the accuracy, the fairness, and the utility of new principal evaluation systems.

This brief provides policymakers an initial review of school climate surveys that have psychometric testing available for review. Most of the surveys included in this brief have not been developed for the express purpose of evaluating school principals, but they have been validated for research or program evaluation purposes. After reviewing this brief, we encourage policymakers to ask climate survey developers and vendors for information on using the surveys for principal evaluation purposes.

This brief identified 11 school climate surveys that displayed publicly available evidence of psychometric rigor (see Table 1). We believe that it is likely that additional school climate surveys have strong psychometric properties, but we were unable to locate evidence about these surveys through our Internet search or efforts to correspond with authors.

Our review suggests that school climate can be measured through surveys of school staff, parents, and students, and each group provides a different perspective on a school. Six surveys were intended for use with school staff only, two were written for school staff and students, and three were written for staff, students, and parents. Some climate surveys have versions for certain types of respondents (e.g., SCI-R) that have been validated for use with all types of respondents, while other surveys have been validated for only one type of respondent (e.g., school staff or students). When selecting school climate surveys for principal evaluation or other purposes, it is important to consider the validity of use for different populations and the cost—in terms of the time required for respondents to complete the survey—necessary to gather accurate information about the school and weigh cost against the potential utility of gathering multiple perspectives on school climate.

The surveys included in this brief also vary on brevity and the constructs measured. At a minimum, they all measure relational supports and organizational supports, with BCSCS (an adapted version of the longer California School Climate Survey) being a good example of a survey designed to be brief. Other surveys measure a host of additional constructs and subconstructs and could take up to 60-90 minutes to complete, examples being the CALL survey or the Essential Supports for School Improvement survey. Although they can be time consuming, they gather more information on additional constructs. For example, the Essential Supports for School Improvement survey measures dimensions such as parent-community ties, student-centered learning climate, and curriculum alignment. As determined by the financial resources available, and the goals, nature, and comprehensiveness of the proposed assessment, administrators will need to decide which stakeholder groups should be surveyed and what constructs should be measured and then choose a survey that meets their needs.
Conclusion

AIR has produced this brief in response to state and school district requests for information about the validity and the reliability of existing, publicly available school climate surveys for use as a measure of principal performance. High-quality principal evaluation systems should be technically sound and logically tied to the standards and the purposes driving the evaluation system design.

Using valid, reliable, and feasible school climate surveys as one measure of principal practice can provide evaluators a more holistic depiction of principal practice. Engaging in a time-consuming and potentially high-stakes principal performance evaluation without first choosing a scientifically sound measure can be a waste of valuable time and limited financial resources. If an ineffective or an inappropriate tool is used to measure broad-based school climate constructs for assessment purposes, misleading findings can lead to an inaccurate evaluation system and, ultimately, wrong decisions.

This brief reviews technical information on 11 school climate surveys that met the minimum criteria for inclusion in the sample as a starting point for identifying viable measures of principal performance. We emphasize that this is a starting point for selection. Policymakers are encouraged to contact survey vendors or technical experts to conduct an in-depth review of school climate surveys and specifically review surveys for

- Financial cost, particularly costs associated with survey analysis and feedback provision.
- Training and support for implementation to ensure reliability.
- Alignment with evaluation purposes, principal effectiveness definitions, and professional standards.

In addition, policymakers should raise questions with vendors about the applicability of climate surveys for elementary, middle, and high schools and procedures for assuring respondent anonymity (a method of ensuring that the survey respondents can respond honestly) and case study or other information about the use of climate surveys for principal feedback. Finally, we encourage policymakers to network with other states or school districts implementing school climate surveys for principal evaluation to learn more about using survey data for summative and formative evaluation purposes.
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Author(s)</th>
<th>Approach</th>
<th>Time Required</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSC-SCAI</td>
<td>Developed by Shindler et al. (2003); published by ASSC</td>
<td>Separate surveys for use with faculty, students (elementary, secondary, and high school), and parents in an individual or a group setting. The various versions range from 30 to 79 items, and all versions address 8 dimensions (physical environment, teacher interactions, student interactions, leadership and decisions, discipline and management, learning and assessment, attitude and culture, and community).</td>
<td>20 minutes</td>
<td>Content validity is established via literature review and theoretical support. Construct validity is established by substantial correlations among the 8 dimensions. Predictive validity is evident by being able to predict student achievement reasonably well based on the survey score.</td>
<td>The subscale reliability coefficients range from .73 to .96. The overall reliability for the scale is .97.</td>
</tr>
<tr>
<td>B-CSCS</td>
<td>You et al. (2009)</td>
<td>A 15-item survey that measures 2 dimensions: relational supports and organizational supports (adapted from the California School Climate Survey). Administered to all school staff (teachers, administrators, and others) to assess school climate.</td>
<td>7–10 minutes (not reported; time inferred by the number of items)</td>
<td>Content validity is based on review of school climate literature and 19 staff school climate surveys. Construct validity is established through confirmatory factor analysis.</td>
<td>The survey contains two subscales: (1) organizational supports: The internal consistency of the subscales ranges from .84 to .86 for teachers and from .79 to .81 for administrator versions across elementary, middle, and high schools. The average subscale reliability for teachers is .85, and the average for administrators is .80. (2) relational supports: The internal consistency of the subscales ranges from .91 to .93 for teacher and administrator versions across elementary, middle, and high schools. The average subscale reliability for teachers is .85, and the average for administrators is .80.</td>
</tr>
<tr>
<td>Instrument</td>
<td>Author(s)</td>
<td>Approach</td>
<td>Time Required</td>
<td>Validity</td>
<td>Reliability*</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>----------</td>
<td>---------------</td>
<td>----------</td>
<td>--------------</td>
</tr>
<tr>
<td>CALL</td>
<td>Halverson et al. (2010)</td>
<td>Principals, teachers, and other staff can take the survey. The principal version has 95 items, and the teacher version has 123 items. The survey focuses on the distribution of leadership in schools, specifically in middle and high schools. The survey contains 5 dimensions (maintaining a schoolwide focus on learning, assessing teaching and learning, collaboratively focusing schoolwide on problems of teaching and learning, acquiring and allocating resources, and maintaining safe and effective learning environment).</td>
<td>45–60 minutes to complete (not reported; time inferred by the number of items)</td>
<td>Content validity is established via extensive review of other measures of school leadership as well as expert review by researchers and practitioners. Construct validity is established by using a Rasch model-based factor analysis.</td>
<td>The Rasch reliability coefficients range from .62 to .87 for five domains. Overall reliability for the survey is .95.</td>
</tr>
<tr>
<td>CSCIb</td>
<td>Developed by CSEE (now the National School Climate Center) in 2004</td>
<td>This 64-item inventory is organized around four school climate dimensions: safety, relationships, teaching and learning, and the environment. It has separate forms for students, school personnel, and parents. Validation for the student version is complete, but validation still needs to be completed for the school personnel and the parent versions.</td>
<td>15–20 minutes</td>
<td>Content validity is established through an extensive literature search and workshops on item development that included teachers, principals, superintendents, and school-based mental health workers. Construct validity is established via confirmatory factor analysis. Convergent validity is established via significant correlations with other measures of nonacademic risk, academic performance, and graduation rates. Analysis of variance, multivariate analysis of variance, and hierarchical linear models show that the subscales and overall scale sufficiently discriminate among schools.</td>
<td>The overall reliability for the scale is .94 for elementary schools and .95 for middle and high schools.</td>
</tr>
<tr>
<td>Instrument</td>
<td>Author(s)</td>
<td>Approach</td>
<td>Time Required</td>
<td>Validity</td>
<td>Reliability&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CGPL Survey</td>
<td>Search Institute (2006)</td>
<td>Focus on psychosocial and learning environment as experienced by students and staff. There are 11 dimensions (55 items) in the student survey and 17 dimensions (76 items) in the staff survey. There are 3 categories of dimensions: relationships, organizational attributes, and personal development.</td>
<td>Not provided (estimated as 30 minutes for the student survey and 40 minutes for the faculty/staff survey)</td>
<td>Content validity is established by theoretical and empirical work in educational psychology. Discriminant and convergent validity is established via correlations with other scales in the survey. Construct validity is established via factor analysis. Predictive validity is established through significant correlations between dimensions and student grade point average.</td>
<td>Low to moderate: The student survey dimension reliabilities range from .60 to .85. (Note: Most dimensions have 4 or fewer items, which hampers reliability.) The test-retest reliabilities for dimensions range from .61 to .87. Low to high: The faculty/staff survey dimension reliabilities range from .68 to .85. (Note: Most dimensions have 5 or fewer items, which hampers reliability.) The test-retest reliabilities for dimensions range from .65 to .90.</td>
</tr>
<tr>
<td>CEEA</td>
<td>Institute for Excellence and Ethics</td>
<td>Separate student, faculty/staff, and parent surveys. Student (75 items) and faculty/staff surveys (105 items) include 3 constructs (with additional subconstructs): safe, supportive, and engaging climate; culture of excellence; and ethics (separate student behaviors and faculty/staff practices constructs for culture of excellence and ethics). Faculty/staff survey also includes a fourth construct for professional community and school/home partnership. The parent survey (54 items) includes 5 constructs; parents’ perceptions of school culture, school engaging parents, parents engaging with school, learning at home/promoting excellence, and parenting/promoting ethics.</td>
<td>Not provided (estimated as 35-40 minutes for the student survey and 45-50 minutes for the faculty/staff survey)</td>
<td>Practitioners and research experts established content validity through reviews of items. Discriminant and convergent validity is established via correlations with external scales.</td>
<td>Moderate to high: student survey construct reliabilities range from .85 to .91. Moderate to high: faculty/staff survey construct reliabilities range from .84 to .93. Moderate to high (apart from one construct with low reliability): parent survey construct reliabilities range from .64 to .91 with a high school/middle school sample and from .68 to .92 for an elementary school sample. (Note: The construct reliabilities that are .64 and .68 are for a construct with only 5 items.)</td>
</tr>
<tr>
<td>Instrument</td>
<td>Author(s)</td>
<td>Approach</td>
<td>Time Required</td>
<td>Validity</td>
<td>Reliability&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>----------</td>
<td>---------------</td>
<td>----------</td>
<td>------------------------</td>
</tr>
<tr>
<td>The Essential Supports for School Improvement</td>
<td>Developed by Sebring et al. (2006); published by the Consortium on Chicago School Research</td>
<td>A framework was developed with accompanying items to better understand the supports that need to be in place in a school to increase learning. Five supports were developed: leadership, parent-community ties, professional capacity, student-centered learning climate, and curriculum alignment. To measure the 5 essential supports, 283 survey items (205 items for teachers, 70 items for students, and 8 items for principals) can be used. The exact number of items to be used depends on the constructs that the surveyor intends to measure.</td>
<td>About 1.5 hours to complete all teacher items and about 30 minutes for the student items (not reported; time inferred by the number of items)</td>
<td>Content validity is established via a literature search, expert review, and the testing of survey items over several years. Construct validity is established via Rasch analyses and relating the 5 supports to indicators of student performance.</td>
<td>The Rasch individual reliabilities for the subscales range from .64 to .92. The Rasch school reliabilities for the subscales range from .55 to .88. The average subscale reliability for individuals is .78; for schools, it is .67.</td>
</tr>
<tr>
<td>ISC-T</td>
<td>Brand et al. (2008)</td>
<td>This 29-item assessment addresses 6 dimensions (peer sensitivity, disruptiveness, teacher-pupil interactions, achievement orientation, support for cultural pluralism, safety problems). It collects information about teachers’ view of school climate.</td>
<td>15–20 minutes per survey (not reported; time inferred by the number of items)</td>
<td>There is extensive validation across three studies. Content validity is based on extensive literature review of existing measures (including ISC-S [student version]) of educational climate as well as literature on how well adolescents adapt to learning environments. Construct validity is established through exploratory and confirmatory factor analysis using diverse samples of schools. Convergent and divergent validity is established with moderate relationships between ISC-S and ISC-T.</td>
<td>The alpha coefficients for subscales range from .57 to .86, with most subscale reliabilities greater than .76. The alpha coefficient for entire survey is .89.</td>
</tr>
<tr>
<td>Instrument</td>
<td>Author(s)</td>
<td>Approach</td>
<td>Time Required</td>
<td>Validity</td>
<td>Reliability(^a)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>TMCI-SF</td>
<td>Sink &amp; Spencer (2007)</td>
<td>This 24-item instrument has 5 factors: satisfaction, competitiveness, difficulty, peer relations, and SCI (school counselor impact or influence)</td>
<td>10–15 minutes (not reported; time inferred by the number of items)</td>
<td>Content validity is established via a literature search. &lt;br&gt;Construct validity is established through exploratory and confirmatory factor analysis.</td>
<td>The subscale alpha coefficients range from .57 to .88, with most subscale reliabilities greater than or equal to .73. The average subscale reliability is .77.</td>
</tr>
<tr>
<td>SCI-R(^b)</td>
<td>Developed by Butler and Alberg (originally published in 1989 and revised in 2002); research on instrument presented in Butler and Rakow (1995); published by CREP</td>
<td>This 49-item assessment that addresses 7 dimensions: order, leadership, environment, involvement, instruction, expectations, and collaboration. It is administered to faculty only.</td>
<td>20 minutes</td>
<td>Content validity is based on a review of literature on factors associated with effective schools and organizational climates. &lt;br&gt;Construct validity is confirmed during the development of the survey in that the items and scales can discriminate among schools.</td>
<td>The internal reliability coefficients for the 7 subscales range from .73 to .84. The average subscale reliability is .76.</td>
</tr>
<tr>
<td>TELL Survey</td>
<td>Research on measure reported by Swanlund (2011); published by the New Teacher Center</td>
<td>The revised survey measures 8 constructs (time, facilities and resources, community support and involvement, managing student conduct, teacher leadership, school leadership, professional development, and instructional practices and support). Each construct contains numerous items; states can add, delete, or revise items to align the survey with their specific context.</td>
<td>About 20 minutes</td>
<td>Content validity is established through an extensive literature review, item-measure correlations, and the fit of the items to model expectations. &lt;br&gt;Validity is established via Rasch analysis.</td>
<td>The Rasch reliability coefficients for subscales range from .80 to .98. The average subscale reliability is .91.</td>
</tr>
</tbody>
</table>

\(^a\)All average reliability coefficients in this table were calculated using Fisher’s z transformation. \(^b\)Denotes an instrument that was identified by Gangi (2010) as being content and psychometrically sound. \(^c\)BCSCS is included in lieu of the complete version of the survey because the complete version has not been validated as of 2011.
References


