Tools

How to use the School Survey of Practices Associated with High Performance

Phyllis Weinstock
Futoshi Yumoto
Yasuyo Abe
IMPAQ International

Coby Meyers
Yinmei Wan
American Institutes
for Research

Overview

This report describes and explains how to use the School Survey of Practices Associated with High Performance, which measures the degree to which schools are engaging in practices associated with high performance. State education departments and school districts can use the survey results to identify and describe school practices associated with high performance, compare practices across school subgroups, target schools for specific interventions, and design interventions. The survey, designed to be taken by teachers and school administrators, measures practices in the domains of effective leadership, strong curriculum, professional development, school culture, and ongoing data use for school improvement. The survey has undergone psychometric validation. The report also includes the survey and describes its development and validation.

IES NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL MIDWEST
Regional Educational Laboratory
At American Institutes for Research

Summary

Regional Educational Laboratory Midwest, through its School Turnaround Research Alliance, developed a survey that state education departments and school districts can use to measure the degree to which schools are engaging in practices associated with high performance. Survey results can be used to identify and describe school practices associated with high performance, compare practices across school subgroups, target schools for specific interventions, and design interventions. The survey, designed to be taken by teachers and school administrators (including principals and assistant principals), measures practices in the domains of effective leadership, strong curriculum, professional development, school culture, and ongoing data use for school improvement. The survey has undergone psychometric validation. This report describes the survey and how to use it, as well as the survey development and validation process.

Contents

What is the School Survey of Practices Associated with High Performance?

The School Survey of Practices Associated with High Performance was developed by Regional Educational Laboratory (REL) Midwest to measure five domains that the research literature suggests are associated with high-performing schools: effective leadership, strong curriculum, professional development, school culture, and ongoing data use for school improvement. The survey elicits information from teachers and school administrators (including principals and assistant principals) about practices in their schools in these five domains.

Why was this survey developed?

To recognize school performance or to learn from local school practices and policies, many states and school districts have identified schools that perform better than expected given the populations they serve ("beating-the-odds" schools). Such schools hold promise for policymakers, researchers, and practitioners alike, suggesting that academic success can be achieved in challenging school environments. This survey emerged from REL Midwest's work with the Michigan Department of Education, research staff from Macomb and Ottawa intermediate school districts,[1] and Learning Forward Michigan. REL Midwest worked with these partners as part of the School Turnaround Research Alliance,[2] focusing on identifying practices, policies, and organizational factors that may be related to high-performing schools.

In supporting the Michigan Department of Education, REL Midwest and the School Turnaround Research Alliance undertook a series of research and development tasks. One task was the development and validation of a teacher and school administrator survey that measures school practices in domains shown to be associated with high performance. Although many surveys designed to measure school effectiveness have been validated, the Michigan Department of Education sought to produce a survey with particular relevance to high-performing schools that were beating the odds. The survey can be used to assess the degree to which schools are engaging in practices associated with high performance. The survey can also help identify areas in which schools (or a subset of schools) need to improve.

The School Survey of Practices Associated with High Performance was developed to measure five domains that are associated with high-performing schools: effective leadership, strong curriculum, professional development, school culture, and ongoing data use for school improvement

How was this survey developed?

Survey development, which included evaluation of the reliability and validity of survey domains (see box 1 for definitions of terms), encompassed seven steps:
- Identifying key domains of practices and policies in which high-performing schools engage.
- Searching for existing surveys that measured the key domains and supporting constructs of the developed conceptual model (see appendix A).
- Selecting items that measure the identified conceptual model constructs.
- Reviewing and revising the survey instrument.
- Conducting cognitive interviews with teachers and principals to identify problems with item wording, comprehension, and recall (see appendix B).
- Conducting a small-scale pilot of the survey and psychometric analysis of pilot survey data.
- Validating the survey with a larger sample.

What literature was consulted to inform survey development?

To produce a survey with measures of school practices that have conceptual and psychometric underpinnings, the study team's initial literature review focused on research on effective schools and school improvement. The accumulated work of Bryk, Bender-Sebring, Allensworth, Luppescu, and Easton (2010) served as the foundational piece. Over 15 years Bryk and colleagues at the University of Chicago Consortium on Chicago School Research administered annual teacher and principal (and student) surveys (http://ccsr. uchicago.edu/surveys/documentation) that eventually led to the identification of five essential domains of high-performing schools:

- *Effective leaders.* The principal works with teachers to implement a clear and strategic vision for school success.
- *Collaborative teachers.* The staff is committed to the school, receives strong professional development, and works together to improve the school.
- *Involved families.* The entire school staff builds strong relationships with families and communities to support learning.
- *Supportive environment.* The school is safe and orderly. Teachers have high expectations for students. Students are supported by their teachers and peers.
- *Ambitious instruction.* Classes are academically demanding and engage students by emphasizing the application of knowledge.

A subsequent literature review by the study team limited to the term "beating the odd(s)" unearthed additional research articles, reports, and policy briefs. Details about the literature review are included in appendix C.[3] Since most publications are derived from case studies or correlational and longitudinal studies, readers should be cautioned that many of these studies do not provide causal evidence of successful strategies for beating the odds.[4] Moreover, the study team identified five domains in that literature associated with high

To produce a survey with measures of school practices that have conceptual and psychometric underpinnings, the initial literature review focused on research on effective schools and school improvement; a subsequent literature review by the study team limited to the term "beating the odd(s)" unearthed additional research

performance (similar to the ones Bryk and colleagues identified), and developed the survey to help measure these domains:

- *Effective leadership.* The administrator's ability to establish a shared mission and associated goals and to provide instructional guidance appears to be a key factor in driving high-performing schools (Beesley & Barley, 2006; Waits et al., 2006).
- *Strong curriculum (with a focus on literacy).* Structured curricular goals at the school and classroom levels, with an emphasis on literacy, are associated with high-performing schools (Mid-continent Research for Education and Learning, 2005; Waits et al., 2006).
- *Professional development.* High-performing schools often afford teachers opportunities to both collaborate and attend meaningful professional development trainings (Langer, 2000; Mid-continent Research for Education and Learning, 2005).
- *School culture.* An orderly environment that encourages parental involvement while emphasizing high academic standards is a consistent feature of schools that are beating the odds (Beesley & Barley, 2006; Grusenmeyer, Fifield, Murphy, Niam, & Qian, 2010; Mid-continent Research for Education and Learning, 2005; Socias, Dunn, Parrish, Muraki, & Woods, 2007).
- *Ongoing data use for school improvement.* Staff, including principals, often review student data and make decisions based on the patterns observed (Grusenmeyer et al., 2010; Southall, 2008).

Why administer this survey?

This survey can be used broadly by states and districts to identify and describe school practices associated with high performance, compare practices across school subgroups, target schools for specific interventions, and design interventions. This survey is best viewed as one instrument with which to collect teacher and administrator perception data on school practices in domains associated with high performance. State or district leaders may then supplement findings with interviews, documents, or observations to better understand school practices and inform policy or other changes. Interventions can also be designed at the state or district level and adapted to target the needs of individual schools. Based on survey findings about specific patterns of practice in each school, interventions might focus on, for example, targeted strategies for teacher or administrator professional development, curriculum and standards alignment, student support, parent involvement, or other aspects of school culture. The survey can be administered annually, or at other regular frequencies as needed, to identify trends and changes in school practices.

Statewide administration of the survey can be used to obtain information on differences in patterns of practice across school subgroups, including both high- and low-performing schools.

The survey also can help create profiles of schools to be disseminated as models for other schools and districts. However, the survey, by itself, cannot conclusively identify the factors that have enabled schools to perform at a high level. The survey findings may highlight areas of practice at each school that could then be further examined and described through interviews and observations.

This survey can be used to identify and describe school practices associated with high performance, compare practices across school subgroups, target schools for specific interventions, and design interventions

3

How to administer this survey

School, district, or state education agency personnel can administer the survey to teachers and administrators online or in paper-and-pencil format. Survey completion requires approximately 20 minutes. The minimum sample size per school needed for generating school-level estimates is 13 respondents for a reliability of 0.60 and 20 respondents for a reliability of 0.70.[5] School practices may differ across subgroups, such as elementary, middle, and high schools; urban and rural schools; or schools with varying student demographic and language backgrounds. The sample for a statewide survey should be stratified to ensure representation of the subgroups of greatest interest for policy reasons.[6]

How to use this survey

Survey users may administer the full survey or select the domains of greatest policy interest

The survey is made up of a series of items, followed by an agreement scale, with respondents asked to select one of four response options: strongly disagree, disagree, agree, or strongly agree.[7] Following the conceptual framework that emerged from the research conducted and expert advice solicited, survey items were organized into five domains: effective leadership, strong curriculum, professional development, school culture, and ongoing data use for school improvement (table 1). The domain measures have undergone psychometric validity testing (see appendix D). Survey users may administer the full survey or select the domains of greatest policy interest. Although subdomains (for example, "organizational direction" and "collaborative leadership" are subdomains under the domain "effective leadership"; see the full survey provided later in this report) are currently identified as parts of the domains, the validation of the survey focused on the larger domain levels. Therefore, it is recommended that each domain be administered in full (that is, with all items that measure the domain) and that survey analysis focus on the domain level.

Descriptive analysis of survey results can be used to explore practices in and across schools. Descriptive statistics used for school profiles and comparisons by domain also can be analyzed by means, medians, standard deviations, and minimum and maximum values; cross-tabulations can show relationships in patterns across several domains or subdomains (for example, the subdomain "organizational leadership" within the domain "effective leadership"). Multiple regression analysis may be used to analyze the contribution of the domain ratings to school performance. Hypothesis testing—using *t*-tests or one-way analysis of variance—can assess hypotheses regarding the domain score differences between sets of schools.[8] Using stratified samples, as described earlier, makes it possible to focus the analysis on subgroups of special interest.

Table 1. Survey domains by item numbers

Domain	Number of items	Item numbers
Effective leadership	18	12a–d, 13a–g, 14a–g
Strong curriculum	19	15a–f, 16a–h, 17a–e
Professional development	13	18a–i, 19a–d
School culture	42	20a–f, 21a–c, 22a–f, 23a–m, 24a–f, 25a–d, 26a–d
Ongoing data use for school improvement	7	27a–g

Source: Authors' compilation.

Instructions on how to calculate a simple composite for a domain from the raw scores of all items in that domain as well as instructions on how to convert raw scores to scaled scores (Rasch ability scores) are included in appendix E.

The survey is presented in the following section as it was administered to a subset of Michigan schools. The survey introduction may be adapted as needed to explain the specific uses of the results and describe privacy protections. Additional survey items identifying the school or district or requesting further information on respondents' positions or backgrounds may be added to the initial section as needed. It was originally administered online but can be administered in paper-and-pencil format.

Descriptive analysis of survey results can be used to explore practices in and across schools

School Survey of Practices Associated with High Performance

Thank you for your participation in this research. We know that your time is valuable, and we greatly appreciate your willingness to complete this survey. The survey was developed to better understand the policies and practices of schools. The questions in this survey focus on your perceptions of your school, and can be analyzed by item and domain, with domain scores computed as the sum or average of item ratings within the domain or as scaled scores obtained through statistical analysis (see appendix E). Domains include the following:

- Effective leadership
- Strong curriculum
- Professional development
- School culture
- Ongoing data use for school improvement

The results of this survey will be used for research and program development purposes. Your responses will be kept completely confidential. Individual responses will not be provided to your principal, your district, or any other party. All survey results will be reported only in statistical summaries that ensure that individuals cannot be identified. This survey will take approximately 20 minutes to complete. If you have any questions or would like more information about the study, you may contact [*insert appropriate contact information*].

1. How do you classify your position at THIS school? [Mark only one response.]
   - ☐ Regular full-time teacher (in any of grades prekindergarten–12 or comparable ungraded levels) [Skip questions 3, 5]
   - ☐ Regular part-time teacher (in any of grades prekindergarten–12 or comparable ungraded levels) [Skip questions 3, 5]
   - ☐ Principal [Skip questions 2, 4, 6, 7]
   - ☐ Assistant principal [Skip questions 2, 4, 6, 7]
   - ☐ Other administrator [Skip questions 2, 4, 6, 7]

2. In what year did you begin teaching at THIS school? _____

3. In what year did you begin working as an administrator at THIS school? _____

4. [TEACHERS ONLY] In what year did you FIRST begin teaching, either full time or part time? _____

5. [ADMINISTRATORS ONLY] In what year did you begin working as a school administrator? _____

6. Do you currently teach students in any of these grades at THIS school? [Select all that apply.]

- ☐ Prekindergarten
- ☐ Kindergarten
- ☐ 1st grade
- ☐ 2nd grade
- ☐ 3rd grade
- ☐ 4th grade
- ☐ 5th grade
- ☐ 6th grade
- ☐ 7th grade
- ☐ 8th grade
- ☐ 9th grade
- ☐ 10th grade
- ☐ 11th grade
- ☐ 12th grade
- ☐ Ungraded

7. This school year, what is your MAIN teaching assignment field at THIS school? (Your main assignment is the field in which you teach the most classes.) [Select all that apply.]

- ☐ Early Childhood or Prekindergarten, General
- ☐ Elementary Education, General
- ☐ Arts or Music
- ☐ English, Reading, or Language Arts
- ☐ English as a Second Language
- ☐ Foreign Languages
- ☐ Health Education
- ☐ Math
- ☐ Natural Sciences
- ☐ Social Sciences or History
- ☐ Career or Technical Education
- ☐ Computer Science
- ☐ Special Education
- ☐ Other

8. Do you have a bachelor's degree?   Yes ☐   No ☐

9. If yes, in what year did you receive your bachelor's degree? _____

10. Do you have a master's degree?   Yes ☐   No ☐

11. If yes, in what year did you receive your master's degree? _____

Effective leadership

Organizational direction

12. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. School administrators make clear the educational goals of the school.	1	2	3	4
b. School administrators maintain high professional expectations for self, faculty, and school.	1	2	3	4
c. School administrators help the faculty develop high professional expectations of themselves.	1	2	3	4
d. School administrators communicate to teachers the directions the school's programs need to take for academic improvement.	1	2	3	4

Collaborative leadership

13. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Administrators, teachers, and staff work together effectively to achieve school goals.	1	2	3	4
b. Teachers can freely provide input and express concerns to administrators.	1	2	3	4
c. The school provides opportunities for parents to participate in important decisions about their children's education (e.g., scheduling, homework, discipline).	1	2	3	4
d. The school ensures teachers have a major role in decisions about curriculum development.	1	2	3	4
e. The school provides opportunities for teachers to plan and make school decisions about professional development and curriculum.	1	2	3	4
f. Teachers have needed instructional resources to teach effectively.	1	2	3	4
g. The school provides regular opportunities for all stakeholders to review the school's vision and purpose.	1	2	3	4

Instructional leadership

14. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. The principal clearly defines or helps teachers understand standards for instructional practices.	1	2	3	4
b. The principal observes teachers teaching.	1	2	3	4
c. The principal attends teacher planning meetings.	1	2	3	4
d. The principal makes suggestions to improve teachers' classroom management.	1	2	3	4
e. The principal gives teachers specific ideas for how to improve instruction.	1	2	3	4
f. The principal empowers teachers to make decisions that improve teaching and learning.	1	2	3	4
g. The principal promotes the diagnosis of individual student learning needs.	1	2	3	4

Strong curriculum (with focus on literacy)

Curriculum, instruction, and assessment aligned with standards

15. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Our staff demonstrates an understanding of state learning standards for reading.	1	2	3	4
b. District or school-level common assessments are used to inform instruction.	1	2	3	4
c. The reading curriculum is aligned with the state learning standards.	1	2	3	4
d. This school uses assessments aligned to standards and curriculum.	1	2	3	4
e. This school uses curriculum that is relevant and meaningful.	1	2	3	4
f. Most teachers integrate literacy concepts into their teaching.	1	2	3	4

Culture of literacy instructional practices

16. Based on your experience, to what extent do you disagree or agree that the following activities are currently practiced throughout your school, across the curriculum? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Most teachers use effective instructional practices in support of developing student literacy and comprehension of course content.	1	2	3	4
b. Most teachers provide personalized support to each student to improve literacy based on assessed needs.	1	2	3	4
c. Most teachers create literacy-rich environments with books, journals, and research texts to support content learning.	1	2	3	4
d. Most teachers effectively use instruction with small groups to improve student learning and comprehension of course content.	1	2	3	4
e. Most teachers effectively model how to use a variety of literacy/learning strategies for all students.	1	2	3	4
f. Most teachers effectively use a variety of literacy strategies that support learning of specific content texts for all students.	1	2	3	4
g. Most teachers regularly use vocabulary development strategies to support student learning.	1	2	3	4
h. Most teachers regularly use strategies to support the reading/writing connection.	1	2	3	4

Culture of literacy intervention to improve student achievement

17. Based on your experience, to what extent do you disagree or agree that the following activities are currently practiced at your school? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Administrators and teachers develop instructional plans to meet literacy instructional needs of struggling students.	1	2	3	4
b. Intervention is highly prescriptive toward improving identified literacy deficits of individuals.	1	2	3	4
c. Highly skilled teachers work with struggling/striving readers.	1	2	3	4
d. Teachers use literacy strategies to support struggling/striving readers' learning of content/subject area texts.	1	2	3	4
e. The school has a plan to improve literacy that supports strategies ranging from intervention for struggling readers to expanding the reading power of all students.	1	2	3	4

Professional development

Focused professional development

18. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Objective data are used to guide building-directed professional development.	1	2	3	4
b. The training I have been to in this district helps me do my job better.	1	2	3	4
c. This school has one or more professional learning communities (a consistent, collaborative learning opportunity for teachers) focused on improving student learning.	1	2	3	4
d. This school's teachers engage in professional development activities to learn and apply reading skills and strategies.	1	2	3	4
e. This school's teachers engage in professional development activities to learn and apply math skills and strategies.	1	2	3	4
f. Teachers in this school are provided with training to collaborate on improving student learning.	1	2	3	4
g. Our teachers engage in classroom-based professional development activities (e.g., peer coaching) that focus on improving instruction.	1	2	3	4
h. We have opportunities to learn effective teaching strategies for the cultures represented in our school.	1	2	3	4
i. We are provided training to support a culturally responsive learning environment.	1	2	3	4

Individual professional development opportunities

19. To what extent do you disagree or agree with the following statements about professional development over the last academic year? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. My professional development has been sustained and coherently focused, rather than short term and unrelated.	1	2	3	4
b. My professional development has included enough time to think carefully about, try, and evaluate new ideas.	1	2	3	4
c. My professional development has been closely connected to my school's improvement plan.	1	2	3	4
d. My professional development has included opportunities to work productively with colleagues in my school.	1	2	3	4

School culture

High academic standards

20. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Students respect others who get good grades.	1	2	3	4
b. Students try hard to improve on previous work.	1	2	3	4
c. Students seek extra work so they can get good grades.	1	2	3	4
d. The school sets high standards for academic performance.	1	2	3	4
e. Students in this school can achieve the goals that have been set for them.	1	2	3	4
f. Academic achievement is recognized and acknowledged by the school.	1	2	3	4

Goal clarity

21. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. School improvement goals are well understood in my school by most teachers and staff.	1	2	3	4
b. The process to achieve school improvement goals is well understood in my school by most teachers and staff.	1	2	3	4
c. School improvement goals give me a sense of direction and purpose for my work.	1	2	3	4

Professional teacher behavior

22. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Most teachers in this school respect the professional competence of their colleagues.	1	2	3	4
b. Most teachers in this school "go the extra mile" with their students.	1	2	3	4
c. Most teachers in this school exercise professional judgment.	1	2	3	4
d. Most teachers in this school accomplish their jobs with enthusiasm.	1	2	3	4
e. Most teachers in this school are committed to helping their students.	1	2	3	4
f. Most teachers in this school help students on their own time.	1	2	3	4

Professional community

23. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Teachers in our school share a similar set of values, beliefs, and attitudes related to teaching and learning.	1	2	3	4
b. In our school we have high expectations for all students.	1	2	3	4
c. Our student assessment practices reflect our curriculum standards.	1	2	3	4
d. Most teachers in the school support the principal in enforcing school rules.	1	2	3	4
e. Most teachers in this school feel responsible for helping each other improve their instruction.	1	2	3	4
f. Most teachers in this school take responsibility for improving the school outside their own class.	1	2	3	4
g. Most teachers in this school help maintain discipline in the entire school, not just their classroom.	1	2	3	4
h. Most teachers in this school observe each other teaching.	1	2	3	4
i. Colleagues provide me with meaningful feedback on my performance.	1	2	3	4
j. Most teachers in this school exchange suggestions for curriculum materials with colleagues.	1	2	3	4
k. Most teachers in this school try to develop new curriculum or lesson plans together.	1	2	3	4
l. Most teachers in this school have conversations with colleagues about managing classroom behavior.	1	2	3	4
m. Most teachers in this school have conversations with colleagues about what helps students learn best.	1	2	3	4

Parent and community involvement

24. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. This school encourages parent involvement.	1	2	3	4
b. Our teachers effectively communicate student progress to parents.	1	2	3	4
c. For important decisions, we collaborate with parents and the community.	1	2	3	4
d. This school communicates effectively with families of all cultures.	1	2	3	4
e. The curriculum we teach reflects the cultures of the community we serve.	1	2	3	4
f. This school has activities to celebrate the cultures of its community.	1	2	3	4

Staff collegiality

25. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. School staff members work well together.	1	2	3	4
b. School staff members are open to feedback regarding their instruction from other staff members.	1	2	3	4
c. I feel comfortable sharing my ideas with other staff members.	1	2	3	4
d. When needed, I can get help and support from other school staff members.	1	2	3	4

School support of innovation

26. Based on your experience, to what extent do you disagree or agree with the following statements? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Leaders support innovation in teaching.	1	2	3	4
b. Most teachers in the school are continually learning and seeking new ideas.	1	2	3	4
c. The principal is interested in innovation and new ideas.	1	2	3	4
d. In my school, we systematically consider new and better ways of doing things.	1	2	3	4

Frequent monitoring of teaching and learning

27. Based on your experience, to what extent do you disagree or agree with the following statements about your school? [Mark only one response.]	Strongly Disagree	Disagree	Agree	Strongly Agree
a. Student assessment results (from either classroom or district assessments) are used to identify student needs and appropriate instructional intervention.	1	2	3	4
b. Struggling students receive early intervention and remediation to acquire skills.	1	2	3	4
c. The administration monitors the effectiveness of instructional interventions.	1	2	3	4
d. School staff reflect upon instructional practice to inform our conversations about improvement.	1	2	3	4
e. Staff are frequently informed about our performance with evidence from observations, student progress, or other data.	1	2	3	4
f. The administration uses data to make recommendations regarding learning programs.	1	2	3	4
g. The administration uses data to assess learning equity for different populations.	1	2	3	4

[END OF SURVEY]

Appendix A. Surveys reviewed in developing the School Survey of Practices Associated with High Performance

The following surveys were located during the search for surveys relevant to development of the School Survey of Practices Associated with High Performance. The 30 surveys that were reviewed are listed here. An asterisk indicates the 10 surveys from which domains were adopted for the School Survey of Practices Associated with High Performance.

Alliance for the Study of School Climate Secondary School Climate Assessment Instrument

Assesses school climate based on eight subfactors: physical appearance, faculty relations, student interactions, leadership and decisionmaking, discipline and management environment, learning instruction and assessment, attitude and culture, and community relations. Respondents include school staff, students, parents, and administrators.
- Alliance for the Study of School Climate. (2011). *Assessment.* Los Angeles, CA: Author.
- Alliance for the Study of School Climate. (2011). *Examining the reliability and validity of the ASSC/WASSC School Climate Assessment Instrument (SCAI).* Los Angeles, CA: Author.

*Audit of Principal Effectiveness

Measures teachers' perceptions of principal effectiveness on nine factors in three domains: organizational development, organizational environment, and educational program.
- Valentine, J. W., & Bowman, M. L. (1986). *Audit of principal effectiveness: A user's technical manual.* Columbia, MO: Authors. http://eric.ed.gov/?id=ED281319

Comprehensive Assessment of School Environments: National Association of Secondary School Principals Student Satisfaction Survey and School Climate Survey

A battery of surveys that measure perceptions of school climate on 10 subscales and satisfaction on 8 subscales. Administered to students, teachers, and parents.
- Lunenberg, F. C. (2011). Comprehensive Assessment of School Environments (CASE): An underused framework for measuring school climate. *National Forum of Educational Administration and Supervision Journal, 29*(4), 1–8.
- McNeal, C. C., & Bishop, H. (1993, November). *A comparative assessment of school environments by delinquent and nondelinquent children: Implications for public school leaders in Alabama.* Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA.

Culture of Excellence and Ethics Assessment Faculty/Staff Survey v4.5

Developed by the Institute for Excellence and Ethics from items on its Collective Responsibility for Excellence and Ethics surveys. The faculty and staff version of the survey measures staff perceptions of school climate and culture on domains related to student safety, engagement, and professional community.
- Khmelkov, V. T., & Davidson, M. L. (2011). *Culture of Excellence & Ethics Assessment Student and Faculty/Staff Survey psychometric data: High school sample.* Lafayette, NY: Institute for Excellence & Ethics.

*Educational Effectiveness Survey

Provides school and district leaders with data related to the characteristics of high-performing schools. Respondents include staff, parents, students, and district and school leadership.

- Center for Educational Effectiveness. (2010). EES v9.0 Research and Development History. In *Educational Effectiveness Survey v9.0 Site Report for West Hills Elem.* (p. 44). Redmond, WA: Author.

Faculty Trust Survey

Measures collective perceptions of school faculty trust in colleagues, the principal, students, and parents. Administered to teachers.

- Hoy, W. K., Smith, P. A., & Sweetland, S. R. (2003). The development of the Organizational Climate Index for high schools: Its measure and relationship to faculty trust. *High School Journal, 86*(2), 38–49.

*Literacy Capacity Survey

Survey for principals and school staff to determine schools' strengths and needs prior to initiating schoolwide programs to improve student literacy.

- National Association of Secondary School Principals. (2005). *Creating a culture of literacy: A guide for middle and high school principals.* Reston, VA: Author. http://eric.ed.gov/?id=ED496862

National School Climate Center Comprehensive School Climate Inventory, Version 3

Developed by the National School Climate Center to understand respondents' perceptions of their schools' socioecological environment on seven domains of school climate. Administered to elementary, middle, and high school students; school staff; and parents.

- Guo, P., Choe, J., & Higgins-D'Alessandro, A. (2011). *Report of construct validity and internal consistency findings for the Comprehensive School Climate Inventory.* New York, NY: Fordham University. Retrieved January 2, 2014, from http://www.schoolclimate.org/climate/documents/Fordham_Univ_CSCI_development_review_2011.pdf.

Ohio State Teacher Efficacy Scale

Measures teachers' perceptions of their efficacy on three scales: instructional strategies, classroom management, and student engagement. Developed by authors in collaboration with students.

- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805.

Organizational Climate Description Questionnaire

Measures three domains of the openness of principal–teacher interactions (supportive, directive, and respective) and three domains of the openness of teacher–teacher interactions (collegial, committed, and disengaged). Administered to teachers.

- Hoy, W. K., Hannum, J., & Tschannen-Moran, M. (1998). Organizational climate and student achievement: A parsimonious and longitudinal view. *Journal of High School Leadership, 8*(4), 336–359.

*Organizational Climate Index

Measures four aspects of school organizational climate: the relationship between the school and the community (institutional vulnerability); the relationship between the principal and teachers (collegial leadership); the relationship among teachers (professional teacher behavior); and teacher, parental, and principal press for achievement (achievement press).

- Hoy, W. K., Smith, P. A., & Sweetland, S. R. (2003). The development of the Organizational Climate Index for high schools: Its measure and relationship to faculty trust. *High School Journal, 86*(2), 38–49.

*Principal Data-Driven Decision-Making Index

Measures principals' use of data-driven decisionmaking. It used the 2002 Educational Leadership Constituent Council/National Council for the Accreditation of Teacher Education program standards as a framework for the survey items and asked about data-driven decisionmaking on four leadership domains: school vision, school instruction, school organizational operation and moral perspective, and collaborative partnerships and larger-context politics. The survey items asked principals to rate their use of data, with data defined as being from four sources: student test scores; demographics, including attendance and graduation rates; teachers', students', administrators', and parents' perceptions of the learning environment; and school programs and instructional strategies.

- Luo, M. (2008). Structural equation modeling for high school principals' data-driven decision making: An analysis of information use environments. *Educational Administration Quarterly, 44*(5), 603–634. Retrieved April 17, 2014, from http://www.emporia.edu/teach/ncate/documents/DDDMarticleinEAQ.pdf.

School Climate Inventory

A survey of school staff perceptions of school climate based on seven domains: order, leadership, environment, involvement, instruction, expectations, and collaboration.

- ReAllen, L., Lowther, D. L., Strahl, J. D., & Slawson, D. (2006). West Orange Collaborative STARK Program 2001–2006 evaluation report. Memphis, TN: Center for Research in Educational Policy. Retrieved April 1, 2014, from http://www.memphis.edu/crep/pdfs/west_orange_students_and_teacher_accessing_real-time_knowledge_program.pdf.

Student Connection Survey

Measures four social and emotional conditions for learning: students are safe, students are challenged, students are supported, and students are socially capable. Administered to middle and high school students.

- American Institutes for Research. (2007). *2007 Student Connection Survey*. Washington, DC: Author. Retrieved March 9, 2014, from http://www.air.org/sites/default/files/downloads/report/CFL_Sample_Score_Report_1690_northside_learning_center_high.pdf.
- Osher, D., Kendziora, K., & Chinen, M. (2008). *Student connection research: Final narrative report to the Spencer Foundation*. Washington, DC: American Institutes for Research. Retrieved April 2, 2014, from http://www.air.org/files/Spencer_final_report_3_31_08.pdf.

*Survey of Chicago Public Schools

Measures teachers' (and principals') perceptions of five essential elements of effective schools: effective leaders, collaborative teachers, involved families, a supportive environment, and ambitious instruction.

- Consortium on Chicago School Research at the University of Chicago. (2007). *Elementary school teacher edition*. Chicago, IL: University of Chicago.
- Consortium on Chicago School Research at the University of Chicago. (2007). *High school teacher edition*. Chicago, IL: University of Chicago.
- Consortium on Chicago School Research at the University of Chicago. (2007). *Principal survey form*. Chicago, IL: University of Chicago.

*Survey of School Policies and Practices

Measures teachers' perceptions of four components selected to describe what differentiates high-performing and low-performing high-need schools: school environment, leadership, professional community, and instruction.

- Apthorp, H., Barley, Z., Englert, K., Gamache, L., Lauer, P., Van Buhler, B., et al. (2004). *McREL's study of academic success in high needs schools: Mid-point progress and measurement viability*. Denver, CO: Mid-continent Research for Education and Learning.
- Wilkerson, S. B., Shannon, L. C., Styers, M. K., & Grant, B. J. (2012). *A study of the effectiveness of a school improvement intervention (Success in Sight): Final report* (NCEE No. 2012–4014). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED530416

Other surveys developed or adapted for specific studies

- Cannata, M., McCrory, R., Sykes, G., Anagnostopoulos, D., & Frank, K. A. (2010). Exploring the influence of National Board Certified Teachers in their schools and beyond. *Educational Administration Quarterly, 46*(4), 463–490.
  - Survey measures perceived teacher leadership activities and teacher influence over schoolwide policy and inclination toward leadership. Items were drawn from the U.S. Department of Education's Schools and Staffing Survey and

previous literature on National Board Certified Teacher leadership activities. The survey was administered to elementary school teachers in two schools in different U.S. regions.

- Daly, A. J., & Chrispeels, J. H. (2008). A question of trust: Predictive conditions for adaptive and technical leadership in educational contexts. *Leadership and Policy in Schools 7*, 30–63.
  - Survey asks teachers and administrators to rate their leadership and trust behaviors as well as the district office's leadership and trust behaviors. District office administrators are asked to rate their leadership and trust behaviors along with those of the average site administrator. The survey measures 11 domains of leadership (culture, order, research-based practices, curriculum and instruction, recognition, involvement, advocacy, empowerment, change, adaptive awareness, and adaptive approaches) and 8 domains of trust (benevolence, respect, communication, openness, integrity, reliability, competence, and risk).

- Eilers, A. M., & Camacho, A. (2007). School culture change in the making: Leadership factors that matter. *Urban Education, 42*(6), 616–637.
  - Annual districtwide survey of teachers on collaborative leadership, evidence-based practices, and communities of practice.

- Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and effect on student achievement. *American Educational Research Journal, 37*(2), 479–507.
  - Survey measures individual teachers' beliefs about the collective school faculty's ability to positively influence student achievement. Adapted by the authors from an earlier teacher efficacy scale: Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 569–582.

- Goddard, Y. L., Neumerski, C. M., Goddard, R. D., Salloum, S. J., & Berebitsky, D. (2010). A multilevel exploratory study of the relationship between teachers' perceptions of principals' instructional support and group norms for instruction in elementary schools. *Elementary School Journal, 111*(2), 336–357.
  - Teacher survey administered to public elementary school teachers in Michigan. Survey developers created a measure of principal instructional leadership and a measure of perceived schoolwide focus on differentiated instruction.

- *Griffith, J. (2003). Schools as organizational models: Implications for examining school effectiveness. *Elementary School Journal, 104*(1), 29–47.
  - Survey developed to study schools as organizational models. Assesses five domains of the school environment: the physical, academic, social, and management environments as well as school–community partnerships. Respondents include school staff and students.

- Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.
  - Scales measuring instructional practices and collaborative leadership developed using a state education agency survey of elementary school teachers, parents, and students. The instructional practice scale consists of four subscales: focus on classroom teaching, quality of support for student learning, professional capacity of the school's teaching staff, and focus on sustained learning improvement. The collaborative leadership scale comprises three aspects of leadership: governance that empowers others and encourages broad participation and responsibility, collaboration in school improvement decisions, and broad participation in evaluating the school's academic development.

- *Heppen, J., Faria, A., Thomsen, K., Sawyer, K., Townsend, M., Kutner, M., et al. (2010). *Using data to improve instruction in the Great City Schools: Key dimensions of practice*. Washington, DC: Council of the Great City Schools. http://eric. ed.gov/?id=ED536737
  - Teacher survey developed to measure teachers' data-use practices in four domains: context, supports for data use, working with data, and instructional responses.

- Jason, M. H. (2001). Principals' self-perceptions of influence and the meaning they ascribe to their leadership roles. *Research for Educational Reform, 6*(2), 34–49.
  - Survey developed to measure the self-perceptions of elementary school principals about the influence and the meaning they derive from their instructional leadership roles. It asks principals to assess their influence as an instructional leader in five areas: school culture, promoting a climate conducive to teaching and learning, enhancing professional development of staff, developing and implementing instructional programs, and obtaining parental involvement and support. It also asks them to assess how much meaning they derived from their instructional leadership actions. The survey was developed specifically for a study in a large urban school district.

- Klecker, B. M., & Pollock, M. A. (2005). Congruency of research-based literacy instruction in high and low performing schools. *Reading Improvement, 42*(3), 149–157.
  - Survey developed for a study on the use of research-based strategies among middle and high school teachers in Kentucky. Survey items asked teachers to rate how much they used each of 20 research-based strategies for teaching reading across the curriculum. The strategies were drawn from the National Institute of Child Health and Human Development's meta-analysis of reading research in 2000, which identified strategies that had a statistically significant positive effect on reading comprehension across grade levels.

- *Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national U.S. survey. *School Effectiveness and School Improvement, 21*(3), 315–336.
  - National teacher survey developed for a project funded by the Wallace Foundation. Respondents were from 180 schools nested within 45 districts nested, in turn, within nine states. The survey includes five scaled variables: focused instruction, teacher's professional community, shared leadership, instructional leadership, and trust. The survey reflects the authors' analytic framework, which assumes that both principal–teacher relationships (indicated by trust, instructional leadership, and perceptions of shared leadership) and teacher–teacher relationships (indicated by professional community) will affect classroom practice—particularly focused instruction—which should, in turn, affect student learning.

- Mattar, D. (2012). Instructional leadership in Lebanese public schools. *Educational Management Administration and Leadership, 40*(4), 509–531.
  - Survey measures principals' instructional leadership style in Lebanese public schools by asking teachers to rate their principals' behaviors. The survey includes two factors: one describing principals' climate-related functions and the other describing principals' technological functions.

- Ware, H., & Kitsantas, A. (2007). Teacher and collective efficacy beliefs as predictors of professional commitment. *Journal of Educational Research, 100*(5), 303–310.
  - The developers selected survey items from the 1999–2000 *Schools and Staffing Survey* and developed two teacher efficacy scales (teacher efficacy to enlist administrative direction and teacher efficacy for classroom management) and a collective efficacy scale and teacher professional commitment scale.

- Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., & Eissa, N. (2009). *Evaluation of the DC Opportunity Scholarship Program: Impacts after three years* (NCEE No. 2009–4050). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED504783
  - School satisfaction index based on dimensions such as safety and academic quality.

Appendix B. Cognitive interviews

The study team developed a cognitive interview protocol to elicit respondent reactions and suggestions regarding construct meaning and relevance, item wording, response scales, and survey practicality for self-administration by working educators. The protocol included think-aloud questions designed to elicit respondents' interpretations and associations with questionnaire items, and a series of direct questions about the survey length, clarity, and alignment of the constructs with respondents' school practices. The protocol also included requests for suggested revisions to improve the survey.

For cognitive interviewing, schools were purposively sampled from one intermediate school district with a representative on the School Turnaround Research Alliance who facilitated initial entry into the schools. The schools were selected based on a range of factors, including school level (that is, one elementary, middle, and high school), various demographic and poverty makeups (that is, race/ethnicity, eligibility for the federal school lunch program, and locale), and rewards school status (which includes high-performing schools that ranked in the 95th–99th percentiles in performance, the 5 percent of schools with the highest rates of improvement, and beating-the-odds schools that are outperforming similar schools, given select risk factors to student achievement). In total the study team conducted cognitive interviews with two teachers and one principal at each of three schools, for nine interviews total.

To obtain some variation, the cognitive interviews were conducted at one elementary school, one junior high school, and one high school. Teacher respondents included a kindergarten teacher, a grade 5 teacher, a middle school math teacher, a middle school English language arts teacher, a high school social studies teacher, and a high school math teacher/department chair. All respondents completed the full survey in preparation for the interview. The interviewer asked each respondent detailed probes and think-aloud questions about items from one or two of the five constructs. Thus, at a minimum, the interviewer received feedback on one complete survey per school. Across the schools all items in the survey received feedback from at least one principal and two teachers (see the table below). The sampling decisions were made to limit cognitive interviews to one hour, an agreement made with the School Turnaround Research Alliance representative who facilitated the cognitive interviews. Thus, the study team ensured receiving holistic feedback on the survey by aggregating cognitive interview data at each school and across educator roles (for example, principals). Interviews were audio-recorded and then professionally transcribed.

The study team synthesized the interview data and implications for possible revisions. The study team read each interview transcript and marked each instance in which a respondent reported difficulty interpreting a question, term, or response category; reported difficulty in trying to formulate an answer; commented on the relevance of a question or scale to his or her specific school or classroom context; or suggested changes to survey items or response scales. The researchers then grouped comments based on their pertinence to specific items, scales, or the survey overall. Next, the study team reviewed each group of comments and flagged items, terms, or scales for which any of the following applied: respondents had differing interpretations of terms and statements; respondents' interpretations were not consistent with the intended domains; respondents reported difficulty with items or response options; items caused confusion or were deemed by respondents to be inapplicable in their schools; or respondents suggested changes. Finally, the study team

grouped comments on the survey overall to identify concerns and suggestions that were not specific to scales or items.

Cognitive interview structure

Domain	Elementary school	Middle school	High school
Effective leadership	Principal	Math teacher	Social studies teacher
Strong curriculum (with a focus on literacy)	Kindergarten teacher	Principal	Math teacher
Professional development	Kindergarten teacher	Principal	Math teacher
School culture	Grade 5 teacher	English teacher	Principal
Ongoing data use for school improvement	Grade 5 teacher	English teacher	Principal

Based on the cognitive interview findings, the study team eliminated several problematic or redundant items and scales, reworded several items, and modified response categories for some items. Because respondents reported difficulty with frequency response categories (never, sometimes, often, always), the study team dropped several items with these response categories and altered some items so the same response categories could be used throughout the survey. Respondents noted that the effective leadership domain items focused almost entirely on centralized leadership and suggested adding measures on distributed leadership inclusive of other school stakeholders. The study team researched additional surveys and identified a validated scale on collaborative leadership that was added to the survey under the effective leadership domain. Overall, based on cognitive reviewers' input on the survey length, and on specific problematic items and redundancies, the study team constructed the pilot survey, which consisted of 105 items.

Appendix C. Related literature on schools that beat the odds

As previously mentioned, the conceptual model that initially guided this survey development project was based on Bryk et al.'s (2010) framework of five domains essential to school improvement, which broadly incorporates school effectiveness, school improvement, beating the odds, and school turnaround. This framework of essential supports guided the study team's search of the literature for relevant constructs. This broad initial scan of the school improvement literature was subsequently narrowed to research that focused on beating-the-odds schools. Using the term "beating the odd(s)," the study team searched ERIC, JSTOR, Wilson Abstracts, and Google for research studies and policy briefs based on research studies (for which original reports were not located). A review of the articles, reports, and policy briefs showed that the evidence base for practices associated with beating the odds relies heavily on case study research. Nevertheless, the review identified five domains of practice that provided a framework for the survey development. This appendix lists 32 references from both the broad search of school improvement literature and the narrower search of the literature focusing on beating-the-odds schools.

American Institutes for Research. (2010). *Identifying potentially successful approaches to turning around chronically low-performing schools: Final report on Study I and plans for Study II.* Washington, DC: Author.

Ascher, C., & Maguire, C. (2007). *Beating the odds: How thirteen NYC schools bring low-performing ninth graders to timely graduation and college enrollment.* Providence, RI: Annenberg Institute for School Reform at Brown University. http://eric.ed.gov/?id=ED522199

Baldwin, L., Coney, F., Fardig, D., & Thomas, R. (1993). *School effectiveness questionnaire.* San Antonio, TX: The Psychological Corporation.

Beesley, A., & Barley, Z. (2006). *Rural schools that beat the odds: Four case studies.* Denver, CO: Mid-continent Research for Education and Learning. Retrieved April 19, 2014, from http://beltonscience.pbworks.com/f/Rural_Beat_the_Odds.pdf.

Bryk, A. S., Bender-Sebring, P., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago.* Chicago, IL: University of Chicago Press.

Carter, S. C. (2000). *No excuses: Lessons from 21 high-performing, high-poverty schools.* Washington, DC: Heritage Foundation. http://eric.ed.gov/?id=ED440170

Chenoweth, K. (2007). *It's being done: Academic success in unexpected schools.* Cambridge, MA: Harvard Education Press.

Chenoweth, K. (2009). *How it's being done: Urgent lessons from unexpected schools.* Cambridge, MA: Harvard Education Press.

Connell, N. (1999). *Beating the odds: High-achieving elementary schools in high-poverty schools.* New York, NY: Education Priorities Panel.

Cudeiro, A., Palumbo, J., & Nelsen, J. (2005). *Six schools that make a difference.* Sacramento, CA: Association of California School Administrators.

Detgen, A., Yamashita, M., Davis, B., & Wraight, S. (2011). *State policies and procedures on response to intervention in the Midwest Region* (REL 2011–No. 116). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. http://eric.ed.gov/?id=ED520327

Education Trust. (2005). *Gaining traction, gaining ground: How some high schools accelerate learning for struggling students.* Washington, DC: Author. Retrieved April 27, 2014, from https://edtrust.org/resource/gaining-traction-gaining-ground-how-some-high-schools -accelerate-learning-for-struggling-students/.

Gottfried, M., Stecher, B., Hoover, M., & Cross, A. B. (2011). *Federal and state roles and capacity for improving schools.* Santa Monica, CA: RAND Corporation.

Grusenmeyer, L., Fifield, S., Murphy, A., Nian, Q., & Qian, X. (2010). *Delaware middle schools: Beating the odds.* Newark, DE: Delaware Education Research and Development Center, University of Delaware.

Harris, D. (2007). High-flying schools, student disadvantage, and the logic of NCLB. *American Journal of Education, 113,* 367–394.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., et al. (2008). *Turning around chronically low-performing schools* (IES Practice Guide, NCEE No. 2008–4020). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED501241

Kearney, W., Herrington, D., & Aguilar, D. (2012). Beating the odds: Exploring the 90/90/90 phenomenon. *Equity & Excellence in Education, 45*(2), 239–249.

Langer, J. (2000). *Beating the odds: Teaching middle and high school students to read and write well.* Albany, NY: University at Albany, State University of New York. Retrieved January 19, 2014, from http://www.albany.edu/cela/reports/langer/langerbeating12014. pdf.

Marzano, R. (2003). *What works in schools: Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Michigan Department of Education. (2010). *Identifying Michigan public schools beating the odds.* Lansing, MI: Author.

Mid-continent Research for Education and Learning. (2005). *Final report: High-needs schools—what does it take to beat the odds?* Washington, DC: U.S. Department of Education. http://eric.ed.gov/?id=ED486626

Pérez, M., Anand, P., Speroni, C., Parrish, T., Esra, P., Socías, M., et al. (2007*). Successful California schools in the context of educational adequacy.* Palo Alto, CA: American Institutes for Research. http://eric.ed.gov/?id=ED499199

Reeves, D. (2004). *Accountability in action: A blueprint for learning organizations.* Denver, CO: Advanced Learning Press.

Sebring, P. B., Allensworth, E., Bryk, A., Easton, J. Q., & Luppescu, S. (2006). *The essential supports for school improvement.* Chicago, IL: Consortium on Chicago School Research, University of Chicago. Retrieved May 1, 2014, from http://ccsr.uchicago.edu/sites/default/files/publications/EssentialSupports.pdf.

Snyder, T. D., & Dillow, S. A. (2011). *Digest of education statistics 2010* (NCES No. 2011–015). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Socias, M., Dunn, L., Parrish, T., Muraki, M., & Woods, L. (2007). *California high schools that beat the odds in high school graduation.* Santa Barbara, CA: California Dropout Research Project. Retrieved June 11, 2014, from http://www.hewlett.org/uploads/files/California_High_Schools_That_Beat_the_Odds_in_High_School_Graduation.pdf.

Southall, S. (2008). *Louisiana High Performing, High Poverty Project.* Baton Rouge, LA: Louisiana Department of Education. Retrieved April 10, 2014, from http://static.discoveryeducation.com/feeds/www/media/pdf/Assessment/Louisiana%20VALED%20Full%20Report.pdf.

Taylor, B., Pearson, P., Clark, K., & Walpole, S. (1999). *Beating the odds in teaching all children to read.* Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. Retrieved April 15, 2014, from http://www.ciera.org/library/reports/inquiry-2/2-006/2-006.pdf.

Teddlie, C., & Reynolds, D. (2000). *The international handbook on school effectiveness research.* New York, NY: Falmer Press.

University of Chicago Consortium on Chicago School Research. (2009). *Elementary teacher survey form and code book.* Chicago, IL: Author.

Usher, A. (2011). *Update with 2009–10 data and five-year trends: How many schools have not made adequate yearly progress?* Washington, DC: Center on Education Policy.

Waits, M. J., Campbell, H. E., Gau, R., Jacobs, E., Rex, T., & Hess, R. K. (2006). *Why some schools with Latino children beat the odds…and others don't.* Tempe, AZ: Morrison Institute for Public Policy. http://eric.ed.gov/?id=ED499748

Appendix D. Piloting and psychometric validation of the survey

This appendix on piloting and psychometric validation findings is included for psychometrically inclined practitioners, that is, practitioners with an understanding of item response theory (IRT)/Rasch modeling (Hambleton, Swaminathan, & Rogers, 1991). The IRT model is a latent (unobservable) trait model that creates a measure of a latent trait on the latent continuum from a set of categorical responses. IRT models provide a defined and common metric for both the latent trait (such as beating the odds–related domains of school practice) and survey items (such as survey item difficulty or, in this survey, the likelihood of an item receiving a strong positive rating).

Piloting the survey

The Michigan Department of Education piloted an online survey in fall 2013 at eight schools. A sample target of approximately 100 administrators and teachers combined was established for the pilot survey, based on minimum requirements for the planned psychometric analysis. A sample of this size is sufficient for flagging items with psychometric challenges. Even though there is no clear guideline to determine the minimum sample size required for IRT analysis (Morizot, Ainsworth, & Reise, 2007), a sample of 100 respondents is considered sufficient for a simple Rasch model. A larger sample may be required, depending on response category usage and model fit. Ultimately, a major factor in assessing whether the sample size is sufficient is whether respondents use all response categories.

Michigan initially identified 27 beating-the-odds schools and sent email invitations to the school principals to solicit their school's participation in the pilot. Eight principals responded and provided teacher and administrator email addresses for their school, resulting in 226 potential respondents—administrators and teachers combined. The survey was then sent to all 226 potential respondents; 95 completed the survey, 8 started the survey but did not complete it, and 123 did not start it.[9] Most respondents were full-time teachers (88 percent). The majority of the teachers (67 percent) who completed the survey had been teaching for more than 10 years. All school levels (elementary, middle, and high schools) were represented among the pilot respondents. Most teachers (68 percent) were teaching middle and high school students (grades 6–12).

Psychometric analysis has two main goals: to establish an instrument's measurement structure and to provide evidence of construct validity. The measurement's purpose is to give a value to a quality or a quantity using a common metric or yardstick. The psychometric validation of the survey included elements of both classical test theory (CTT) (Lord & Novick, 1968) and IRT, with a greater emphasis on IRT analysis.[10] CTT is a raw score–based method using average scores from survey items as the outcome measurement or measurement of a latent construct or domain. For example, on the survey, researchers assigned a number to each category (1 for strongly disagree, 2 for disagree, 3 for agree, and 4 for strongly agree) and then summed these numbers as a measure. However, the sum of ordinal numbers cannot establish measurement properties: the difference of ordinal scores, 2 to 1, does not have any inherent meaning. Therefore, this survey's validation process emphasized model-based measurement, or IRT. IRT expands on the CTT analysis by examining additional item properties that affect validity—measurement of the domains of interest—as well as reliability of the survey. IRT provides additional psychometric information, especially at the item level, beyond that provided by CTT, to examine and establish

the construct validity of measures. In addition, IRT parameters are population invariant or independent, while CTT parameters are sample dependent. The Rasch model analysis, a form of IRT analysis, helps ensure that the beating-the-odds measures are scientifically justifiable and meaningful.

The foundation of the psychometric analysis is the conceptual soundness of the constructs and items, as informed by earlier steps in the validation process: the literature review, expert review, and cognitive interviewing. Although the Rasch analysis can identify problem items, decisions about removing, adding, or changing items, constructs, or domains should always be informed by content knowledge in addition to the psychometric findings.

For the CTT, analyses excluded respondents with missing responses on each domain (that is, list-wise deletion) and calculated the following:
- Response rates.
- Basic statistical calculations (mean, standard deviation).
- Calculation of Cronbach's alpha[11] to assess the internal consistency of item scales measuring each domain: a reliability of 0.70 or above is considered minimally sufficient for exploratory analyses (Nunnally, 1978).[12]
- Analysis of intercorrelations among the domains; these should be positive across the domains, but not so high as to indicate that each domain is not capturing a distinct policy and practices domain. The literature does not provide a rule for acceptable levels of intercorrelation, so judgment must be used in reviewing these findings. Intercorrelations are between .51 and .77 (table D1), well under .85, a cutoff that is commonly used.

The IRT analysis involved analysis of items to assess their fit with the Rasch model assumptions (box D1). The Rasch analysis included the following:
- *Flagging items that are outside the expected range of difficulty (ease of endorsement or agreement) or do not follow the expected relationship between respondents' ability (the underlying level of their opinion or the likelihood of endorsing items) and the difficulty of items.* Flagging criteria are explained in box D1 (see also table D1). This analysis was conducted because an item provides limited information about the respondent (that is, it does not discriminate among respondents well) when it is either too difficult or too easy.
- *Examining local independence of items.* This analysis was conducted because a response to one item should not affect responses to other items.
- *Reviewing evidence of potential multidimensionality within each of five domains.* Multidimensionality violates the assumption that the items associated with each domain measure one primary latent concept. Residual analysis was conducted to identify multidimensionality that might threaten the validity of the primary domain and also to identify potential local item dependencies.
- *Testing marginal reliability.* Marginal reliability, similar to internal consistency reliability in CTT, is reported in Rasch analysis. Levels of marginal reliability should be similar to estimates of reliability based on or derived from applying Cronbach's alpha.

Summary statistics and reliability measures based on the pilot survey data also were examined, as well as descriptive statistics based on CTT (table D2). The table provides basic

Box D1. Definitions of flagging criteria

Flagging criteria based on violations of Rasch model assumptions include the following:

Data-model/Infit index indicates whether the item response pattern of respondents is consistent with the Rasch model's assumptions. The Infit index assesses the magnitude of error in how respondents answered each item. The items flagged by this criterion might add excessive random error to the score or might not add information because this index assesses the threat to construct validity. A high Infit value indicates that excessive error (or unexpected responses) exists; the item therefore does not provide maximum information to differentiate among respondents with different levels of ability (this is the case when the item difficulty is close to the respondents' ability levels).

Difficulty is an item parameter on the logit scale estimated in Rasch analysis. It measures the item's underlying difficulty, with higher values indicating greater difficulty. In the context of this survey a more difficult item is less likely to be endorsed in the agree or especially the strongly agree categories (that is, a more difficult item represents a statement that is less easily endorsed or agreed to, even when higher levels of the underlying domain are present). Items provide more information when they are not too difficult or too easy.

Mean ability score in each response category is expected to increase monotonically (consistently), from the lower to higher categories. In this context respondent ability is measured as the average ability level of respondents in each response category, with higher ability associated with more frequent endorsements in the agree and especially the strongly agree categories. A nonmonotonic pattern of mean scores indicates, for example, that the estimated ability associated with a lower response category is higher than the estimated ability associated with a higher response category. This may indicate that respondents had difficulty discriminating between response categories for the item in question, or that the stem question was unclear or misaligned. The comparison of the mean ability scores is meaningful when the number of responses in each category is sufficiently large. For the pilot data, as shown below, many items had very few responses for categories 1 (strongly disagree) and 2 (disagree): in those cases analyses excluded the mean ability score comparisons across these categories.

Point-measure correlation (PTME) is a correlation between the item categories chosen and the Rasch ability estimates. PTME is an index of the individual item's contribution to the Rasch-modeled estimates. PTME is similar to a point-biserial correlation, a correlation between an item and the raw total score, conventionally used in CTT analysis. An item with a high PTME value has high discrimination value: there is a clear relationship between response categories chosen and respondents' ability levels. A lower PTME value indicates a lower discrimination value and a lesser contribution to measure the construct or domain of interest, or excessive measurement errors.

One flagging criterion is based on CTT analysis:

Proportion of responses in each response category is an indicator of category use. A proportion of responses less than 0.01 indicates that the category is underused, whereas a category is overused when the proportion of responses is greater than 0.95.

Table D1. Basic item flagging criteria

Item statistic	Flagging criteria
Difficulty (Rasch difficulty estimates, d, averaged across response categories) on the Likert scale (from 1 = strongly disagree to 4 = strongly agree)	$d > 3.00$ or $d < -3.00$
Proportion (prop) of responses in each response category (from 1 = strongly disagree to 4 = strongly agree)	prop < 0.01 or prop > 0.95
Mean ability score in each response category (from 1 = strongly disagree to 4 = strongly agree)	The mean ability score in a response category is lower than the mean ability score in the next lower category.
Data-model fit, measured by the Infit index (how well the item data fit the Rasch model)	Infit < 0.6 or Infit > 1.4[a]
Item to Rasch ability estimate correlations, or point-measure (PTME) correlations (correlation between item scores and the total score)	PTME < 0.15

a. Wright & Linacre, 1994, p. 370.

Source: Thissen & Wainer, 2001.

information about the pilot results in terms of the mean item score (on a scale from 1 to 4, with 4 being the most positive rating) for each domain. The mean domain score was generally positive, ranging from 2.9 for professional development to 3.3 for strong curriculum (with a score of 3 representing the agree response category). Reliability alphas were high, ranging from 0.88 for data use to 0.96 for school culture.

Descriptive statistics based on IRT/Rasch analyses were examined (table D3). The estimated mean ability measure ranged from 1.14 for professional development to 2.13 for school culture. Marginal reliabilities were high, ranging from 0.83 for data use to 0.93 for school culture.

Indicators of reliability—Cronbach's alpha (Nunnally & Bernstein, 1994)—also were examined (see table D2, corresponding to CTT; Lord & Novick, 1968), as well as the marginal reliability (see table D3, corresponding to IRT). These reliability indicators measure internal consistency (the extent to which items measuring the same construct or domain

Table D2. Summary statistics for the pilot sample, classical test theory measures

Domain	Number of respondents	Number of items	Mean score	Standard deviation	Minimum score[a]	Maximum score	Reliability: alpha
Effective leadership	100	18	3.2	0.42	2.2	4.0	0.91
Strong curriculum	96	20	3.3	0.43	1.8	4.0	0.95
Professional development	96	15	2.9	0.50	1.2	4.0	0.93
School culture	95	44	3.2	0.39	1.3	4.0	0.96
Ongoing data use for school improvement	95	8	3.2	0.44	2.1	4.0	0.88

Note: Responses on incomplete surveys are included. Mean score is computed based on nonmissing items. Four of 100 respondents completed questions only on the leadership domain. One respondent completed three domains (effective leadership, strong curriculum, and professional development) and partially completed the school culture questions.

a. Minimum scores can be 2 or higher if no respondents selected the lowest response category (1).

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

Table D3. Summary statistics for the pilot sample, item response theory/Rasch measures, and reliability

| Domain | Number of respondents | Rasch estimated ability | | | | Marginal reliability |
		Mean	Standard deviation	Minimum	Maximum	
Effective leadership	96	1.73	1.64	−1.57	6.01	0.87
Strong curriculum	91	1.56	2.04	−5.21	7.16	0.88
Professional development	93	1.14	1.94	−5.98	5.51	0.90
School culture	96	2.13	1.79	−4.10	8.53	0.93
Ongoing data use for school improvement	88	1.73	1.64	−4.58	5.69	0.83

Note: Extreme scores are excluded.

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

are intercorrelated). The reliability measures show that both the Cronbach's alpha and the marginal reliability are consistently high for all domains, ranging from 0.88 to 0.96 on the alpha and from 0.83 to 0.93 on the marginal reliability. See Kline (2000) for a discussion of standards for reliability levels. The review of these reliability measures indicates no major threats to the validity of the survey instrument. However, given the model misfit (responses were missing in some categories), the pilot sample was relatively small. Therefore, the analysis results must be viewed with caution.

In addition to the review of the summary statistics of the CTT and IRT/Rasch measures, the study team examined the pair-wise correlation among the five domains. The study team expected the correlation to be positive across the constructs, but not so high as to indicate that they were not capturing distinct policy and practice constructs. Results from correlational analyses demonstrate that the correlation ranged from 0.51 between curriculum and leadership to 0.77 between curriculum and school culture (table D4). The correlation of 0.77 is relatively high; across the school culture and curriculum constructs, several items may be measuring similar school practices related to alignment of assessment and standards and consistency of teaching strategies and expectations. The persistence of high correlation could indicate that item removal might be considered. However, removing items without violating construct validity would require review of item statistics and flags, and potentially a revision of Rasch estimates based on item parameters.

The IRT and CTT analyses were performed independently for each domain. The study team examined criteria to identify items that may potentially threaten the internal validity of the domains (table D5). These criteria, defined in box D1, were selected based on commonly applied standards. If an item falls within these flagging ranges, it may indicate the following: a misaligned item (too difficult or too easy), a category utilization issue (a category with a very low or very high response rate), item categories are not discriminating among respondents as expected, an item has excessive measurement error, or an item is not contributing to measuring a domain of interest. The flagging criteria are applied as a first-round screening for potential problem items.

If an item does not meet one or more of the criteria, it does not automatically mean that the item should be removed or altered in future administrations. Rather, the criteria are

Table D4. Correlation among survey domains

Domain	Effective leadership	Strong curriculum	Professional development	School culture	Ongoing data use for school improvement
Effective leadership	1.00				
Strong curriculum	0.51	1.00			
Professional development	0.65	0.57	1.00		
School culture	0.68	0.77	0.69	1.00	
Ongoing data use for school improvement	0.56	0.60	0.55	0.66	1.00

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

evaluated to flag the items that may warrant further investigation. The pilot sample size was relatively small for a survey of this complexity, in which the items have multiple response options. This makes the parameter estimates less reliable. Therefore, items with multiple flags were re-examined with data from a larger sample (described below).

The flagging criteria are based primarily on the IRT/Rasch modeling analysis but also include a CTT criterion (a proportion of responses). As further explained the Rasch model assumes a specific relationship among the item properties, the respondents' patterns of endorsing items, and the probability of endorsing an item (see box D1). Items are flagged based on violations of Rasch model assumptions that may indicate threats to the validity of the items and introduce measurement error into the overall score. Two of most important flags are Infit >1.4 and PTME < 0.15. These flags represent the most serious threats to construct validity and to the item's ability to differentiate among different respondents (or schools in the original survey administration).

Analyses examined the number of items within each of the five domains flagged by four criteria (see table D5). No items were flagged based on difficulty. In addition to the item statistics described in table D5, the study team also examined two key assumptions of the Rasch model. One assumption is that the items for each domain measure a predominant, single construct of interest. Although all constructs or domains are expected to have some multidimensionality, a concern can arise if a secondary dimension accounts for a large variance—about half of the primary or measurement variance—and exhibits a systematic (rather than random) pattern. This may indicate consideration of a separate construct or domain. Another assumption is that items are locally independent, given the ability level of the respondent; that is, the response to one item should not affect the response to any other items. Local item dependency is a strong residual dependency between two items. Residual analyses were conducted to check for the potential violation of both the unidimensionality and local independence assumptions. Although the analysis found no serious violation of the unidimensionality assumption, it did find evidence of local item dependencies. As indicated, 10 item pairs in each of the first four domains had residual correlations showing possible local dependencies (a residual correlation of less than –0.30 or greater than 0.30; see table D5). These dependencies may be related to item order.

The analysis of the pilot data found no conclusive threats to validity. The survey showed consistently high reliability for all domains, along with adequate separation across domains. There was no evidence of strong multidimensionality in any domains, even in

Table D5. Number of flagged items by criteria

Domain	Total number of items in domain	Flagging criteria				
		Proportion of responses in each category (<0.01 or >0.95)	The mean ability score (lower than the mean ability score in the next lower category)	Item to Rasch ability measure point measure correlation (<0.15)	Infit Index: <0.6 or >1.4	Local item dependencies
Effective leadership	18	3	8	2	2	10 item pairs
Strong curriculum	20	14	4	2	2	10 item pairs
Professional development	15	2	9	2	2	10 item pairs
School culture	44	8	17	3	3	10 item pairs
Ongoing data use for school improvement	8	6	1	1	1	3 item pairs

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

school culture, which includes the largest number of items measuring several arguably disparate aspects of school culture (parent involvement, staff collegiality, academic pressure). The pilot analysis identified 14 items across the five domains that were flagged on the most serious potential threats, usually high Infit and at least one additional criterion (table D6). The study team recommended that these items be re-examined using a larger sample and suggested the following additional areas for further analysis.

The four response categories (1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree) were not always fully used by pilot respondents. In particular, category 1 (strongly disagree) was seldom selected for any item on the survey. This indicates that there might be systematic factors, such as socially desirable effects, driving response patterns. If this issue persists with a large sample, collapsing of categories (such as reducing the categories to the binary agree/disagree options) or rescoring might be considered.

The mean scores for categories 3 and 4 are reversed for many items. It is expected that the mean scores monotonically increase from the lowest to highest item categories. This may indicate that respondents had difficulty differentiating these two categories for these items. If this issue persists with a larger sample, however, collapsing the two categories and reevaluating the key statistics, including Infit and PTME (which can indicate serious threats to reliability and validity), is suggested.

In many cases the residual correlations between adjacent items were high, suggesting dependencies between items—a violation of a key Rasch model assumption. A simple adjustment, such as reordering items (for example, randomizing the order in which survey questions appear), can reduce the effects of such local item dependencies.

Analysis with a larger sample

In fall 2014 the Michigan Department of Education administered the survey online to a sample of 64 schools during a six-week window. A total of 212 administrators and teachers from 34 schools responded to the survey. The study team conducted the same psychometric analysis as that in the pilot analysis using this larger sample. Also, summary statistics

Table D6. Summary of items flagged for re-examination

Item	N	Rasch analysis			Proportion of responses by category				Mean ability score by category			
		Difficulty	Infit index	Point-measure correlation	1	2	3	4	1	2	3	4
Q13c	100	−0.05	1.44[a]	0.45[b]	0.01	0.09	0.68	0.22	−0.01	0.14	0.94[a]	0.54[a]
Q13f	100	0.64	1.63[a]	0.46[b]	0.03	0.19	0.59	0.19	0.03	0.16	1.00[a]	0.42[a]
Q14c	100	0.36	1.36[b]	0.56[b]	0.03	0.18	0.51	0.28	−0.01	0.15	0.73[a]	0.73[a]
Q15b	96	−0.25	1.59[a]	0.58+	0.03	0.10	0.53	0.33	−0.01	0.01	0.52	0.79
Q17c	96	2.16	1.62[a]	0.58[b]	0[a]	0.33	0.44	0.23	[c]	0.18	0.61[a]	0.52[a]
Q18a	96	−0.46	1.55[a]	0.52[b]	0.03	0.14	0.54	0.29	−0.03	−0.02	0.57	0.59
Q18d	96	0.29	1.15	0.59[b]	0[a]	0.15	0.47	0.39	[c]	−0.03	0.21	0.92
Q19e	96	1.51	1.42[a]	0.63[b]	0.10	0.42	0.39	0.09	0.11	0.69	0.64[a]	0.30[a]
Q20a	96	0.02	1.44[a]	0.43[b]	0.01	0.17	0.54	0.28	−0.04	0.23	0.89[a]	0.66[a]
Q20b	96	0.53	1.23	0.49[b]	0.02	0.07	0.70	0.21	−0.04	0.13	1.12[a]	0.54[a]
Q20e	96	0.42	1.51[a]	0.47[b]	0.01	0.31	0.47	0.21	−0.04	0.41	0.87[a]	0.50[a]
Q26c	95	0.07	1.16	0.46[b]	0[a]	0.04	0.42	0.54	[c]	0.02	0.49	1.21
Q27a	95	1.34	1.24[b]	0.62[b]	0[a]	0[a]	0.55	0.45	[c]	[c]	0.10	0.84
Q27e	95	0.59	1.56[a]	0.67	0.05	0.37	0.43	0.15	−0.07	−0.06	0.74[a]	0.32[a]

Note: Categories are as follows: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.

a. Value met the item flagging criteria described in table D5.

b. Item may warrant closer attention even though it is not flagged by criteria in table D5.

c. Value has blank cells (item categories that received no responses).

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

and reliability measures were reviewed based on this larger sample (tables D7 and D8). The CTT analysis showed relatively high levels of construct reliabilities, with Cronbach's alpha ranging from 0.92 to 0.96, which is consistent with findings from the pilot data. The findings from the Rasch analysis also were consistent with the pilot findings in terms of construct reliabilities (with marginal reliability ranging from 0.85 to 0.95). The Rasch analysis with the larger sample showed no evidence of strong multidimensionality or local item dependency. Of the 14 items that were flagged for re-examination based on pilot data,

Table D7. Summary statistics for the larger sample in 2014, classical test theory measures

Domain	Number of respondents	Number of items	Mean score	Standard deviation	Minimum score[b]	Maximum score	Reliability: alpha
Effective leadership	203	17[a]	3.2	0.47	1.4	4.0	0.92
Strong curriculum	193	19	3.4	0.44	2.2	4.0	0.95
Professional development	191	15	3.0	0.58	1.3	4.0	0.94
School culture	182	42	3.3	0.41	2.2	4.0	0.96
Ongoing data use for school improvement	182	7	3.2	0.56	1.1	4.0	0.92

Note: Responses on incomplete surveys are included. Mean score is computed based on nonmissing items.

a. One item that was asked in the pilot was not included in the pilot with this larger sample.

b. Minimum scores can be 2 or higher if no respondents selected the lowest response category (1).

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

Table D8. Summary statistics for the larger sample in 2014, item response theory/Rasch measures, and reliability

| Domain | Number of respondents | Rasch estimated ability | | | | Marginal reliability |
		Mean	Standard deviation	Minimum	Maximum	
Effective leadership	207	2.07	1.90	–2.58	6.67	0.88
Strong curriculum	197	3.59	2.29	–1.18	7.71	0.87
Professional development	192	1.72	2.33	–3.76	7.13	0.91
School culture	189	2.33	1.70	–1.09	8.49	0.95
Ongoing data use for school improvement	182	2.83	2.67	–5.05	7.25	0.85

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

six of them (table D9) continued to have high Infit values from the Rasch analysis, which suggested that these items may pose threats to the validity of the scales. The study team removed these six items from the survey (table D10) and repeated the psychometric analysis; the findings were consistent with previous analyses and did not lower correlations between constructs.

Because the pilot sample and the larger sample came from the same population and the differential item functioning analyses (that is, examination of item difficulty parameter differences by subgroups, using two test administrations) showed consistent item calibrations, the study team combined the two samples and repeated the Rasch analysis to obtain more precise estimates of Rasch model parameters (Howard & Wainer, 1993). The Rasch scores presented in appendix E are based on this combined sample.

Table D9. Summary of removed items

| Original item number | Number of respondents | Rasch analysis | | | Proportion of responses by category | | | | Mean ability score by category | | | |
		Difficulty	Infit index	Point-measure correlation	1	2	3	4	1	2	3	4
Q17c	193	1.70	1.63[a]	0.64[b]	0.04	0.19	0.44	0.33	1.82[a]	1.54[a]	3.08	5.59
Q18a	192	–0.68	1.43[a]	0.62[b]	0.04	0.11	0.46	0.39	–1.49	–0.09	0.92	3.47
Q19e	191	1.49	1.48[a]	0.66	0.12	0.35	0.37	0.17	–0.23	0.68	1.82	5.06
Q20a	189	1.61	1.57[a]	0.49[b]	0.06	0.29	0.43	0.21	0.88	1.45	2.53	3.57
Q20e	189	2.04	1.58[a]	0.38[b]	0.05	0.39	0.44	0.12	2.42[a]	1.67[a]	2.35[a]	4.31
Q27e	184	1.94	1.49[a]	0.73	0.09	0.24	0.43	0.24	0.33	0.73	2.64	6.14

Note: Categories are as follows: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.

a. Value met the item flagging criteria described in table D5.

b. Item may warrant closer attention even though it is not flagged by criteria in table D5.

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

Table D10. Items removed from survey

Original item number	Domain	Subdomain	Item
Q17c	Strong curriculum	Culture of literacy intervention to improve student achievement	Ample tutoring sessions are available to support improved student literacy.
Q18a	Professional development	Focused professional development	My principal (or administrator) talks to me about my professional development.
Q19e	Professional development	Individual professional development opportunities	My professional development has included opportunities to work productively with teachers from other schools.
Q20a	School culture	High academic standards	Parents exert pressure to maintain high standards.
Q20e	School culture	High academic standards	Parents press for improvement of the school.
Q27e	Ongoing data use for school improvement	Frequent monitoring of teaching and learning	Teacher observations of other teachers lead to meaningful change in instructional practice.

Note: Categories are as follows: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.

Source: Authors' analysis based on data from the School Survey of Practices Associated With High Performance (2014).

Appendix E. Calculating domain scores

For Likert response survey data (as in the chosen scale of strongly disagree, disagree, agree, and strongly agree) users often want to combine responses from multiple questions into a composite score and then use the composite score(s) as a variable of interest in analysis. This appendix describes two approaches to generating composite domain scores for users of this survey.

One approach is simply averaging each participant's response values on all domain questions. This approach includes the following: converting or recoding ordinal (Likert) responses to numeric responses, with lower values indicating more negative responses (1 for strongly disagree, 2 for disagree, 3 for agree, and 4 for strongly agree); calculating the average domain score by dividing the sum of raw scores by the number of completed items in the domain;[13] or imputing the missing responses with the average score from responded items. The latter approach, using imputation, is recommended.[14] This approach to generating composite scores is straightforward and easy to understand. The resulting sums or average scores are best used to describe or summarize responses and to analyze relative differences among schools rather than magnitudes of differences.

Another approach is to generate a scaled score for each domain based on results from the Rasch analysis. Scaled scores (Rasch ability scores) rather than sums or averages of raw scores are recommended for use in analyses that require interval-level data: calculations of means, standard deviations, multiple regressions, and analysis of variance. For example, scaled scores could be used as independent variables in a regression to examine whether a given domain score (for example, a leadership score) predicts student performance outcomes. Scaled scores for each of the five survey domains were examined (tables E1–E5). The Rasch ability score precision, or the standard error of measurement, is different for all score points, and also is presented in the tables. Users of this survey can convert raw scores to scaled scores using these tables.[15]

The Rasch model is a probabilistic model that can be examined to establish measurement or scale properties. The Rasch score is defined on an interval-level scale (logit scale). In contrast a simple sum or average of ordinal survey responses, such as a Likert scale, is not based on a defined metric. The distance between two categories on the ordinal or Likert scale is undefined; it is difficult to know whether the difference between strongly disagree and disagree is comparable to the difference between strongly agree and agree. Raw scores are frequently used for Likert-scale responses and can be an adequate approximation to measure the construct or domain of interest. However, the results have much weaker validity than results based on scaled scores and are best used to compare relative differences (for example, rank order) rather than magnitudes of differences. This suggests that teachers at one school might indicate higher levels of effective leadership, for example, than teachers at another school do. But this does not allow for interpretations that attempt to articulate how sizable the difference really is.

Table E1. Rasch ability scores for effective leadership domain

Sum of raw scores	Rasch ability score	Standard error	Sum of raw scores	Rasch ability score	Standard error
≤18	−6.11	1.83	46	−0.26	0.36
19	−4.89	1.01	47	−0.12	0.37
20	−4.17	0.72	48	0.01	0.38
21	−3.74	0.60	49	0.16	0.39
22	−3.43	0.53	50	0.31	0.40
23	−3.18	0.48	51	0.48	0.41
24	−2.97	0.44	52	0.65	0.42
25	−2.78	0.42	53	0.83	0.43
26	−2.62	0.40	54	1.01	0.44
27	−2.46	0.38	55	1.21	0.44
28	−2.32	0.37	56	1.41	0.45
29	−2.19	0.36	57	1.61	0.45
30	−2.06	0.35	58	1.81	0.45
31	−1.94	0.34	59	2.01	0.45
32	−1.83	0.34	60	2.21	0.45
33	−1.71	0.33	61	2.41	0.45
34	−1.60	0.33	62	2.62	0.45
35	−1.50	0.33	63	2.83	0.46
36	−1.39	0.33	64	3.04	0.47
37	−1.28	0.33	65	3.27	0.48
38	−1.17	0.33	66	3.51	0.50
39	−1.07	0.33	67	3.78	0.53
40	−0.96	0.33	68	4.09	0.57
41	−0.85	0.33	69	4.45	0.64
42	−0.74	0.34	70	4.93	0.76
43	−0.62	0.34	71	5.69	1.03
44	−0.50	0.35	72	6.95	1.84
45	−0.38	0.35			

Source: Authors' analysis based on data from the School Survey of Practices Associated with High Performance (2014).

Table E2. Rasch ability scores for strong curriculum domain

Sum of raw scores	Rasch ability score	Standard error	Sum of raw scores	Rasch ability score	Standard error
≤19	–7.56	1.84	48	–0.32	0.42
20	–6.32	1.03	49	–0.15	0.42
21	–5.57	0.75	50	0.04	0.43
22	–5.11	0.63	51	0.23	0.45
23	–4.76	0.56	52	0.44	0.46
24	–4.47	0.51	53	0.65	0.48
25	–4.22	0.48	54	0.89	0.49
26	–4.00	0.46	55	1.14	0.51
27	–3.80	0.44	56	1.41	0.52
28	–3.61	0.43	57	1.68	0.53
29	–3.42	0.42	58	1.96	0.53
30	–3.25	0.41	59	2.23	0.52
31	–3.08	0.41	60	2.49	0.50
32	–2.91	0.41	61	2.74	0.49
33	–2.75	0.40	62	2.97	0.48
34	–2.59	0.40	63	3.20	0.47
35	–2.43	0.40	64	3.41	0.46
36	–2.27	0.40	65	3.62	0.45
37	–2.11	0.40	66	3.82	0.45
38	–1.95	0.40	67	4.03	0.46
39	–1.79	0.40	68	4.24	0.46
40	–1.63	0.40	69	4.47	0.48
41	–1.47	0.40	70	4.70	0.50
42	–1.31	0.40	71	4.96	0.52
43	–1.15	0.40	72	5.26	0.57
44	–0.99	0.40	73	5.61	0.63
45	–0.83	0.40	74	6.08	0.75
46	–0.66	0.41	75	6.84	1.03
47	–0.49	0.41	76	8.08	1.84

Source: Authors' analysis based on data from the School Survey of Practices Associated with High Performance (2014).

Table E3. Rasch ability scores for professional development domain

Sum of raw scores	Rasch ability score	Standard error	Sum of raw scores	Rasch ability score	Standard error
≤13	−6.97	1.85	33	−0.21	0.50
14	−5.70	1.05	34	0.04	0.51
15	−4.91	0.77	35	0.30	0.52
16	−4.41	0.66	36	0.58	0.54
17	−4.02	0.59	37	0.88	0.55
18	−3.69	0.55	38	1.20	0.57
19	−3.40	0.53	39	1.53	0.58
20	−3.14	0.51	40	1.86	0.58
21	−2.89	0.49	41	2.19	0.57
22	−2.65	0.48	42	2.51	0.56
23	−2.43	0.47	43	2.82	0.56
24	−2.20	0.47	44	3.13	0.55
25	−1.99	0.46	45	3.44	0.56
26	−1.77	0.46	46	3.75	0.57
27	−1.56	0.46	47	4.09	0.59
28	−1.34	0.46	48	4.45	0.62
29	−1.13	0.47	49	4.87	0.68
30	−0.91	0.47	50	5.40	0.79
31	−0.68	0.48	51	6.21	1.06
32	−0.45	0.49	52	7.49	1.86

Source: Authors' analysis based on data from the School Survey of Practices Associated with High Performance (2014).

Table E4. Rasch ability scores for school culture domain

Sum of raw scores	Rasch ability score	Standard error	Sum of raw scores	Rasch ability score	Standard error
≤42	−7.78	1.83	87	−1.42	0.24
43	−6.56	1.01	88	−1.36	0.25
44	−5.84	0.73	89	−1.30	0.25
45	−5.41	0.60	90	−1.23	0.25
46	−5.09	0.53	91	−1.17	0.25
47	−4.84	0.48	92	−1.11	0.25
48	−4.63	0.44	93	−1.05	0.25
49	−4.45	0.41	94	−0.99	0.25
50	−4.29	0.39	95	−0.93	0.25
51	−4.15	0.37	96	−0.87	0.25
52	−4.01	0.36	97	−0.80	0.25
53	−3.89	0.34	98	−0.74	0.25
54	−3.78	0.33	99	−0.68	0.25
55	−3.67	0.32	100	−0.61	0.25
56	−3.57	0.32	101	−0.55	0.26
57	−3.47	0.31	102	−0.48	0.26
58	−3.38	0.30	103	−0.42	0.26
59	−3.29	0.30	104	−0.35	0.26
60	−3.20	0.29	105	−0.28	0.26
61	−3.12	0.29	106	−0.22	0.26
62	−3.04	0.28	107	−0.15	0.26
63	−2.96	0.28	108	−0.08	0.27
64	−2.89	0.27	109	−0.01	0.27
65	−2.81	0.27	110	0.07	0.27
66	−2.74	0.27	111	0.14	0.27
67	−2.67	0.26	112	0.21	0.27
68	−2.60	0.26	113	0.29	0.27
69	−2.53	0.26	114	0.36	0.27
70	−2.47	0.26	115	0.44	0.28
71	−2.40	0.26	116	0.51	0.28
72	−2.33	0.25	117	0.59	0.28
73	−2.27	0.25	118	0.67	0.28
74	−2.21	0.25	119	0.75	0.28
75	−2.14	0.25	120	0.83	0.28
76	−2.08	0.25	121	0.91	0.28
77	−2.02	0.25	122	0.99	0.29
78	−1.96	0.25	123	1.07	0.29
79	−1.90	0.25	124	1.15	0.29
80	−1.84	0.25	125	1.23	0.29
81	−1.78	0.25	126	1.32	0.29
82	−1.72	0.25	127	1.40	0.29
83	−1.66	0.25	128	1.48	0.29
84	−1.60	0.24	129	1.57	0.29
85	−1.54	0.24	130	1.65	0.29
86	−1.48	0.24	131	1.74	0.29

(continued)

Table E4. Rasch ability scores for school culture domain *(continued)*

Sum of raw scores	Rasch ability score	Standard error	Sum of raw scores	Rasch ability score	Standard error
132	1.82	0.29	151	3.58	0.33
133	1.91	0.29	152	3.69	0.34
134	1.99	0.29	153	3.81	0.34
135	2.08	0.29	154	3.93	0.35
136	2.16	0.29	155	4.06	0.36
137	2.25	0.29	156	4.19	0.37
138	2.34	0.30	157	4.34	0.38
139	2.43	0.30	158	4.49	0.40
140	2.51	0.30	159	4.65	0.41
141	2.60	0.30	160	4.83	0.43
142	2.69	0.30	161	5.03	0.46
143	2.78	0.30	162	5.25	0.49
144	2.88	0.31	163	5.50	0.52
145	2.97	0.31	164	5.80	0.57
146	3.07	0.31	165	6.17	0.65
147	3.16	0.31	166	6.67	0.77
148	3.26	0.32	167	7.45	1.05
149	3.37	0.32	168	8.73	1.85
150	3.47	0.33			

Source: Authors' analysis based on data from the School Survey of Practices Associated with High Performance (2014).

Table E5. Rasch ability scores for ongoing data use for school improvement domain

Sum of raw scores	Rasch ability score	Standard error	Sum of raw scores	Rasch ability score	Standard error
≤7	−8.00	1.91	20	1.09	0.96
8	−6.57	1.15	21	2.18	1.11
9	−5.57	0.89	22	3.35	1.01
10	−4.87	0.79	23	4.24	0.88
11	−4.28	0.75	24	4.96	0.82
12	−3.73	0.74	25	5.62	0.81
13	−3.19	0.74	26	6.32	0.88
14	−2.63	0.75	27	7.26	1.11
15	−2.06	0.76	28	8.62	1.88
16	−1.49	0.76	25	5.62	0.81
17	−0.92	0.76	26	6.32	0.88
18	−0.33	0.77	27	7.26	1.11
19	0.30	0.83			

Source: Authors' analysis based on data from the School Survey of Practices Associated with High Performance (2014).

Notes

1. An intermediate school district in Michigan is a government agency organized to assist regionally based school districts by providing programs and services.
2. The Beating the Odds Research Alliance was renamed at the beginning of this project. The alliance's new name is the School Turnaround Research Alliance.
3. Appendix C includes 32 research articles, reports, and policy briefs, including the accumulated work of Bryk and colleagues.
4. See Herman et al. (2008) for information regarding the strength of causal evidence required from studies rated by the What Works Clearinghouse, as well as recommendations for turning around chronically low-performing schools.
5. The minimum sample size was estimated based on the pilot data and by incorporating intraclass correlations into the Spearman-Brown prediction formula (Raudenbush & Bryk, 2001; Winer, Brown, & Michels, 1991). The reliability of a school's score depends on the number of questions in a scale, the number of staff members in a school completing the survey, and the extent to which schools naturally vary in the outcome.
6. Stratified sampling is the process of first grouping members of the population into relatively homogeneous subgroups (called strata) and then creating a sample by drawing subsamples from each of those subgroups. The sample size for each subgroup is proportional to the size of the subgroup. Stratification helps ensure that a representative mix of units is selected from the population and aids in ensuring that there is a large enough sample to generate estimates for relevant groups of interest. One purpose of using stratified sampling in education survey research could be to ensure that surveys are administered to a sample of schools that reflect a state's demographics. For instance, researchers may wish to stratify by geographic location to ensure that an adequate mix of schools from different geographic areas (for example, rural, urban, suburban) are selected. See Kish (1965) for a detailed explanation.
7. Survey research indicates that when given a noncommittal or more neutral option, respondents are more likely to choose it (Nardi, 2005). Thus, a forced choice was applied—that is, not offering a neither agree nor disagree response option.
8. Whether applied to a simple random sample or stratified sample, these basic analytic procedures will not differ.
9. The study team's primary goal was to attain about 100 teacher respondents as a minimum criterion for pilot data analysis.
10. The IRT analysis for this study used a Rasch analysis based on a partial-credit model, which is a polytomous extension of a Rasch model (a one-parameter model; Masters, 1982), using Winsteps software.
11. The reliability measure used for the IRT analysis is marginal reliability rather than Cronbach's alpha. Cronbach's alpha is presented here for those interested in CTT analysis
12. What Works Clearinghouse, Procedures and Standards Handbook (version 3.0) indicates alpha of 0.50 as an acceptable level of reliability.
13. Four of 100 respondents analyzed in the initial pilot completed questions only on the effective leadership construct. One respondent completed three domains (effective leadership, strong curriculum, and professional development) and partially completed

the school culture questions. Moreover, this aggregation method treats each item in the domain equally. One may assign more weight to some items and less weight to the others. Moreover, if the number of responses varies across items, one may also want to give more weight to items with larger numbers of responses by calculating an average weighted by the number of responses for each item.

14. For more information on imputing missing data, see Allison (2001).

15. The use of Rasch ability scores in tables E1–E5 requires complete responses or no missing data.

References

Alliance for the Study of School Climate. (2011). *Assessment.* Los Angeles, CA: Author.

Alliance for the Study of School Climate. (2011). *Examining the reliability and validity of the ASSC/WASSC School Climate Assessment Instrument (SCAI).* Los Angeles, CA: Author.

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage Publications, Inc.

American Institutes for Research. (2007). *2007 Student Connection Survey.* Washington, DC: Author. Retrieved March 9, 2014, from http://www.air.org/sites/default/files/downloads /report/CFL_Sample_Score_Report_1690_northside_learning_center_high.pdf.

American Institutes for Research. (2010). *Identifying potentially successful approaches to turning around chronically low-performing schools: Final report on Study I and plans for Study II.* Washington, DC: Author.

Apthorp, H., Barley, Z., Englert, K., Gamache, L., Lauer, P., Van Buhler, B., et al. (2004). *McREL's study of academic success in high needs schools: Mid-point progress and measurement viability.* Denver, CO: Mid-continent Research for Education and Learning.

Ascher, C., & Maguire, C. (2007). *Beating the odds: How thirteen NYC schools bring low-performing ninth graders to timely graduation and college enrollment.* Providence, RI: Annenberg Institute for School Reform at Brown University. http://eric.ed.gov/?id=ED522199

Baldwin, L., Coney, F., Fardig, D., & Thomas, R. (1993). *School effectiveness questionnaire.* San Antonio, TX: The Psychological Corporation.

Beesley, A., & Barley, Z. (2006). *Rural schools that beat the odds: Four case studies.* Denver, CO: McREL. Retrieved April 19, 2014, from http://beltonscience.pbworks.com/f/ Rural_Beat_the_Odds.pdf.

Bryk, A. S., Bender-Sebring, P., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago.* Chicago, IL: University of Chicago Press.

Cannata, M., McCrory, R., Sykes, G., Anagnostopoulos, D., & Frank, K. A. (2010). Exploring the influence of National Board Certified Teachers in their schools and beyond. *Educational Administration Quarterly, 46*(4), 463–490.

Carter, S. C. (2000). *No excuses: Lessons from 21 high-performing, high-poverty schools.* Washington, DC: Heritage Foundation. http://eric.ed.gov/?id=ED440170

Center for Educational Effectiveness. (2010). EES v9.0 Research and Development History. In *Educational Effectiveness Survey v9.0 Site Report for West Hills Elem.* (p. 44). Redmond, WA: Author.

Chenoweth, K. (2007). *It's being done: Academic success in unexpected schools.* Cambridge, MA: Harvard Education Press.

Chenoweth, K. (2009). *How it's being done: Urgent lessons from unexpected schools.* Cambridge, MA: Harvard Education Press.

Connell, N. (1999). *Beating the odds: High-achieving elementary schools in high-poverty neighborhoods.* New York, NY: Education Priorities Panel.

Consortium on Chicago School Research at the University of Chicago. (2007). *Elementary school teacher edition.* Chicago, IL: University of Chicago.

Consortium on Chicago School Research at the University of Chicago. (2007). *High school teacher edition.* Chicago, IL: University of Chicago.

Consortium on Chicago School Research at the University of Chicago. (2007). *Principal survey form.* Chicago, IL: University of Chicago.

Cudeiro, A., Palumbo, J., & Nelsen, J. (2005). *Six schools that make a difference.* Sacramento, CA: Association of California School Administrators.

Daly, A. J., & Chrispeels, J. H. (2008). A question of trust: Predictive conditions for adaptive and technical leadership in educational contexts. *Leadership and Policy in Schools 7,* 30–63.

Detgen, A., Yamashita, M., Davis, B., & Wraight, S. (2011). *State policies and procedures on response to intervention in the Midwest Region* (REL 2011–No. 116). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. http://eric.ed.gov/?id=ED520327

Education Trust. (2005). *Gaining traction, gaining ground: How some high schools accelerate learning for struggling students.* Washington, DC: Author. Retrieved April 27, 2014, from https://edtrust.org/resource/gaining-traction-gaining-ground-how-some-high-schools -accelerate-learning-for-struggling-students/.

Eilers, A. M., & Camacho, A. (2007). School culture change in the making: Leadership factors that matter. *Urban Education, 42*(6), 616–637.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and effect on student achievement. *American Educational Research Journal, 37*(2), 479–507.

Goddard, Y. L., Neumerski, C. M., Goddard, R. D., Salloum, S. J., & Berebitsky, D. (2010). A multilevel exploratory study of the relationship between teachers' perceptions of principals' instructional support and group norms for instruction in elementary schools. *Elementary School Journal, 111*(2), 336–357.

Gottfried, M., Stecher, B., Hoover, M., & Cross, A. B. (2011). *Federal and state roles and capacity for improving schools*. Santa Monica, CA: RAND Corporation.

Griffith, J. (2003). Schools as organizational models: Implications for examining school effectiveness. *Elementary School Journal, 104*(1), 29–47.

Grusenmeyer, L., Fifield, S., Murphy, A., Nian, Q., & Qian, X. (2010). *Delaware middle schools: Beating the odds*. Newark, DE: Delaware Education Research and Development Center, University of Delaware.

Guo, P., Choe, J., & Higgins-D'Alessandro, A. (2011). *Report of construct validity and internal consistency findings for the Comprehensive School Climate Inventory*. New York, NY: Fordham University. Retrieved January 2, 2014, from http://www.schoolclimate.org/climate/documents/Fordham_Univ_CSCI_development_review_2011.pdf.

Hambleton, R. K., Swaminathan, H., & Rogers, H. F. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Press.

Harris, D. (2007). High-flying schools, student disadvantage, and the logic of NCLB. *American Journal of Education, 113*, 367–394.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.

Heppen, J., Faria, A., Thomsen, K., Sawyer, K., Townsend, M., Kutner, M., et al. (2010). *Using data to improve instruction in the Great City Schools: Key dimensions of practice*. Washington, DC: Council of the Great City Schools. http://eric.ed.gov/?id=ED536737

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., et al. (2008). *Turning around chronically low-performing schools* (IES Practice Guide, NCEE No. 2008–4020). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED501241

Howard, P. W., & Wainer, H. (1993). *Differential item functioning*. Mahwah, NJ: Erlbaum.

Hoy, W. K., Hannum, J., & Tschannen-Moran, M. (1998). Organizational climate and student achievement: A parsimonious and longitudinal view. *Journal of High School Leadership, 8*(4), 336–359.

Hoy, W. K., Smith, P. A., & Sweetland, S. R. (2003). The development of the Organizational Climate Index for high schools: Its measure and relationship to faculty trust. *High School Journal, 86*(2), 38–49.

Jason, M. H. (2001). Principals' self-perceptions of influence and the meaning they ascribe to their leadership roles. *Research for Educational Reform, 6*(2), 34–49.

Kearney, W., Herrington, D., & Aguilar, D. (2012). Beating the odds: Exploring the 90/90/90 phenomenon. *Equity & Excellence in Education, 45*(2), 239–249.

Khmelkov, V. T., & Davidson, M. L. (2011). *Culture of Excellence & Ethics Assessment Student and Faculty/Staff Survey psychometric data: High school sample.* Lafayette, NY: Institute for Excellence & Ethics.

Klecker, B. M., & Pollock, M. A. (2005). Congruency of research-based literacy instruction in high and low performing schools. *Reading Improvement, 42*(3), 149–157.

Kish, L. (1965). *Survey sampling.* New York, NY: John Wiley & Sons, Inc.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London, UK: Routledge.

Langer, J. (2000). *Beating the odds: Teaching middle and high school students to read and write well.* Albany, NY: University at Albany, State University of New York. Retrieved January 19, 2014, from http://www.albany.edu/cela/reports/langer/langerbeating12014. pdf.

Lord, F. M., & Novick, M. R. (1968.) *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national U.S. survey. *School* Effectiveness *and School Improvement, 21*(3), 315–336.

Lunenberg, F. C. (2011). Comprehensive Assessment of School Environments (CASE): An underused framework for measuring school climate. *National Forum of Educational Administration and Supervision Journal, 29*(4), 1–8.

Luo, M. (2008). Structural equation modeling for high school principals' data-driven decision making: An analysis of information use environments. *Educational Administration Quarterly, 44*(5), 603–634. Retrieved April 17, 2014, from http://www.emporia.edu/ teach/ncate/documents/DDDMarticleinEAQ.pdf.

Marzano, R. (2003). *What works in schools: Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.

Mattar, D. (2012). Instructional leadership in Lebanese public schools. *Educational Management Administration and Leadership, 40*(4), 509–531.

McNeal, C. C., & Bishop, H. (1993, November). *A comparative assessment of school environments by delinquent and nondelinquent children: Implications for public school leaders in Alabama.* Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA.

Michigan Department of Education. (2010). *Identifying Michigan public schools beating the odds*. Lansing, MI: Author.

Mid-continent Research for Education and Learning. (2005). *Final report: High-needs schools: What does it take to beat the odds?* Aurora, CO: Author. http://eric.ed.gov/?id=ED486626

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 407–423). New York, NY: Guilford.

Nardi, P. M. (2005). *Doing survey research: A guide to quantitative methods.* (2nd ed.). New York, NY: Pearson.

National Association of Secondary School Principals. (2005). *Creating a culture of literacy: A guide for middle and high school principals.* Reston, VA: Author. http://eric.ed.gov/?id=ED496862

Nunnally, J. (1978). *Psychometric theory.* New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Office of Quality Improvement. (2010). *Survey fundamentals: A guide to designing and implementing surveys.* Madison, WI: Author.

Osher, D., Kendziora, K., & Chinen, M. (2008). *Student connection research: Final narrative report to the Spencer Foundation.* Washington, DC: American Institutes for Research. Retrieved April 2, 2014, from http://www.air.org/files/Spencer_final_report_3_31_08.pdf.

Pérez, M., Anand, P., Speroni, C., Parrish, T., Esra, P., Socías, M., et al. (2007). *Successful California schools in the context of educational adequacy.* Palo Alto, CA: American Institutes for Research. http://eric.ed.gov/?id=ED499199

Raudenbush, S. W., & Bryk, A. W. (2001). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

ReAllen, L., Lowther, D. L., Strahl, J. D., & Slawson, D. (2006). West Orange Collaborative STARK Program 2001–2006 evaluation report. Memphis, TN: Center for Research in Educational Policy. Retrieved April 1, 2014, from http://www.memphis.edu/crep/pdfs/west_orange_students_and_teacher_accessing_real-time_knowledge_program.pdf.

Reeves, D. (2004). *Accountability in action: A blueprint for learning organizations.* Denver, CO: Advanced Learning Press.

Sebring, P. B., Allensworth, E., Bryk, A., Easton, J. Q., & Luppescu, S. (2006). *The essential supports for school improvement.* Chicago, IL: Consortium on Chicago School

Research, University of Chicago. Retrieved May 1, 2014, from http://ccsr.uchicago.edu/sites/default/files/publications/EssentialSupports.pdf.

Snyder, T. D., & Dillow, S. A. (2011). *Digest of education statistics 2010* (NCES 2011–015). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Socias, M., Dunn, L., Parrish, T., Muraki, M., & Woods, L. (2007). *California high schools that beat the odds in high school graduation.* Santa Barbara, CA: California Dropout Research Project. Retrieved June 11, 2014, from http://www.hewlett.org/uploads/files/California_High_Schools_That_Beat_the_Odds_in_High_School_Graduation.pdf.

Southall, S. (2008). *Louisiana High Performing, High Poverty Project.* Baton Rouge, LA: Louisiana Department of Education. Retrieved April 10, 2014, from http://static.discoveryeducation.com/feeds/www/media/pdf/Assessment/Louisiana%20VALED%20Full%20Report.pdf.

Taylor, B., Pearson, P., Clark, K., & Walpole, S. (1999). *Beating the odds in teaching all children to read.* Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. Retrieved April 15, 2014, from http://www.ciera.org/library/reports/inquiry-2/2-006/2-006.pdf.

Teddlie, C., & Reynolds, D. (2000). *The international handbook on school effectiveness research.* New York, NY: Falmer Press.

Thissen, D., & Wainer, H. (Eds.). (2001). *Test scoring.* New York: Routledge.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805.

University of Chicago Consortium on Chicago School Research. (2009). *Elementary teacher survey form and code book.* Chicago, IL: Author.

Usher, A. (2011). *Update with 2009–10 data and five-year trends: How many schools have not made adequate yearly progress?* Washington, DC: Center on Education Policy.

Valentine, J. W., & Bowman, M. L. (1986). *Audit of principal effectiveness: A user's technical manual.* Columbia, MO: Authors. http://eric.ed.gov/?id=ED281319

Waits, M. J., Campbell, H. E., Gau, R., Jacobs, E., Rex, T., & Hess, R. K. (2006). *Why some schools with Latino children beat the odds…and others don't.* Tempe, AZ: Morrison Institute for Public Policy. http://eric.ed.gov/?id=ED499748

Ware, H., & Kitsantas, A. (2007). Teacher and collective efficacy beliefs as predictors of professional commitment. *Journal of Educational Research, 100*(5), 303–310.

Wilkerson, S. B., Shannon, L. C., Styers, M. K., & Grant, B. J. (2012). A study of the effectiveness of a school improvement intervention (Success in Sight): Final report (NCEE No. 2012–4014). Washington, DC: U.S. Department of Education, Institute

of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED530416

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design*. New York, NY: McGraw–Hill.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., & Eissa, N. (2009). *Evaluation of the DC Opportunity Scholarship Program: Impacts after three years* (NCEE No. 2009–4050). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED504783

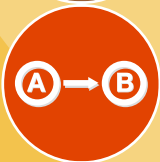Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research