Asking the Right Questions About Using Nonacademic Indicators for School Accountability

By Laura S. Hamilton

March 2021

Key Points

- Policymakers increasingly face questions about whether test-based accountability paints an overly narrow picture of school performance.
- One proposed strategy for reducing reliance on test scores is to incorporate non-test measures—"nonacademic indicators" such as school climate surveys—into accountability systems.
- While including nonacademic indicators could lead to better decision-making and, ultimately, better outcomes for students, rushed implementation of non-validated metrics could carry significant costs.
- Policymakers who wish to implement nonacademic indicators should ask themselves what their goals are in adding nonacademic indicators, whether adding indicators is the only way to achieve those goals, who should be involved in selecting indicators, and how data will actually be used.

Whether and how to hold schools accountable for their performance has been a topic of debate in education circles for decades, and the use of standardized achievement test scores for accountability has been particularly contentious. One proposed strategy for reducing reliance on test scores is to incorporate non-test measures into accountability systems. In fact, doing so is a requirement under the Every Student Succeeds Act (ESSA), the federal legislation that guides state testing and accountability policy.

Additional indicators can reduce the extent to which test scores drive school performance ratings while sending a message to educators and the public about the need for schools to attend to conditions or outcomes beyond test scores. In this report, I discuss a category of measures that I call "nonacademic" indicators and briefly describe ways they can be used to inform decision-making. I then offer a set of questions that policymakers should ask before adopting such measures for accountability purposes.

Understanding the Current School Accountability Context

Research on whether school accountability has been beneficial, harmful, or inconsequential is mixed but suggests that the effects depend on the system's features and the context in which accountability is enacted. Despite this mixed record, a testing backlash has been brewing since the passage of

the No Child Left Behind Act in 2001, which mandated state tests along with consequences for performance on them. This backlash has gained steam in the past five years,[1] with critics representing all parts of the political spectrum.

Recent events have fueled new objections to the uses of assessment for school accountability. Because of COVID-19, states canceled their spring 2020 tests, and many traditional school and classroom assessments had to be scrapped or redesigned. At the same time, COVID-19-related school closures have made it abundantly clear to parents who suddenly became full-time homeschool instructors that schools are responsible for many services and outcomes beyond those measured by standardized achievement tests, such as creating opportunities for social interactions and keeping students motivated.

Meanwhile, the role that schools and other educational institutions play in promoting racial justice and equity has been highlighted by recent events and has reinvigorated long-standing discussions of public education's responsibilities in this area. Black Lives Matter protests created a sense of urgency for schools to address racism, and some scholars and education leaders are increasingly calling on schools to promote civic skills and dispositions including appreciation for diversity and an understanding of sources of racial disparities.[2]

The movement has also amplified the voices of critics of standardized tests who believe these tests propagate racial inequity. The reasons for persistent differences in scores across racial and ethnic groups are a topic of intense debate, with some blaming the tests themselves, despite substantial evidence that these score differences stem from unequal access to learning opportunities and broader societal supports from an early age.[3]

This confluence of events has renewed debates regarding the fundamental purposes and responsibilities of public schools, with potential implications for how we hold schools accountable. At the time of this report, the timing and format of spring 2021 state testing is unknown. Regardless of what form that testing takes, state and local education leaders are likely to face resistance and questions about whether these tests are worthwhile or necessary and, in particular, whether they paint an overly narrow picture of school performance. One way to mitigate this concern while ensuring that parents and others continue to receive information about schools' performance is to incorporate additional, non-test-based indicators into the system.

What Are Indicators?

Throughout this report, I refer to indicators and draw on a recent National Academies of Sciences, Engineering, and Medicine (NASEM) report on indicators of educational equity. The NASEM report defines an indicator as "a measure used to track progress toward objectives or to monitor the health of an economic, environmental, social, or cultural condition over time."[4] The report describes two broad categories of indicators: indicators of student outcomes including achievement and attainment and indicators of access to resources and opportunities, such as effective teaching or adequate school funding. In these categories, the report's authors list several features that characterize high-quality indicators, including comparability across contexts, developmental appropriateness, and scientific soundness of measures.

Most importantly, it is crucial that the indicators align with the system's purposes. An accountability purpose requires indicators that represent outcomes or activities over which schools exercise at least some control. Thus, measures such as childhood exposure to lead in water or paint, which clearly affects opportunity to learn and should be considered as part of any broad effort to reduce educational inequity, might not be suitable for a typical school accountability system because schools generally have little control over that exposure.

What Are "Nonacademic" Indicators, and How Are They Used?

For the purposes of this report, I distinguish between academic and nonacademic indicators while acknowledging some ambiguity in definitions. Academic indicators include achievement test scores and other measures of academic performance such as graduation rates or academic course-taking histories. Nonacademic indicators, by contrast, include those that capture evidence regarding school and student performance in other domains such as

school climate and safety or student social and emotional learning (SEL) competencies that are relevant to schools' curricula. Examples of nonacademic indicators include scores on surveys that measure the quality of student-staff relationships, student engagement, or student motivation.

As with most concepts related to measurement, the distinctions between these definitions are not clear-cut. Graduation, for example, reflects a combination of academic and nonacademic opportunities to learn. Consensus on these definitions is not essential for this report's key points, most of which apply to any type of indicator.

A comprehensive review of how nonacademic indicators are used in state and local accountability systems is beyond the scope of this report. In this section, I briefly describe two examples of how such indicators have been incorporated into broader systems of measurement.

The ESSA Fifth Indicator. The 2015 reauthorization of ESSA was intended to mitigate some perceived problems with the narrow focus on tests under previous authorizations. One way it does so is through the so-called "fifth indicator" in states' ESSA plans. Formally known as the school quality/student success (SQSS) indicator, it enables states to incorporate measures that do not fit in the other four required indicator categories.[5] The measures used for this indicator must meet certain requirements,[6] but states have extensive leeway to select measures that are aligned with their goals, and these measures can be academic or nonacademic. Moreover, the SQSS indicator can be constructed by aggregating across multiple indicators of process or performance.

A 2018 FutureEd analysis of states' ESSA plans[7] identified chronic absenteeism as the most widely used SQSS indicator. Measures of college and career readiness were also prevalent, and the FutureEd authors point out that these indicators—based on measures such as advanced course-taking or postsecondary enrollment rates—generally fall into the academic rather than nonacademic category. A small number of states did adopt nonacademic indicators beyond chronic absenteeism. These include suspension rates (four states) and survey-based measures of climate or engagement (eight states).

A separate analysis showed that the climate and engagement surveys in states' ESSA plans covered a diverse array of topics including bullying, school safety, relationships with other students or adults, and respect for diversity.[8] But, these surveys typically receive low weight (5–10 percent) in the overall accountability index, with the bulk of the weight being placed on academic achievement test scores and graduation rates. While this suggests nonacademic indicators probably do not significantly influence accountability decisions, their inclusion in states' ESSA plans could send a signal about what the state values, and the data they produce might be useful for informing local improvement efforts. Next, I describe an example of how districts have used nonacademic indicators for local improvement.

CORE Districts. The CORE Districts—a consortium of school districts in California—developed an indicator system that includes both academic and nonacademic indicators and is used primarily to support quality improvement at the school level. Using surveys administered annually to students in grades four through 12, the CORE Districts' system provides data on four student SEL competencies and four aspects of school climate and culture.[9]

CORE also gathers climate data through annual surveys of teachers and other school staff as well as parents and other caregivers. These measures are not part of a high-stakes accountability system, but the CORE Districts' experience is relevant to this report because its system is unique in its systematic collection of, and reporting on, both climate and SEL competencies. Schools receive reports with aggregate data, and some districts use dashboards to share data with stakeholders including school-level educators and parents. District leaders have drawn on these data for such purposes as documenting gaps in climate perceptions and SEL competencies among students from different racial and ethnic groups, identifying schools where black and Latino students are producing particularly favorable or unfavorable results, and informing hypotheses regarding the reasons for gaps in academic achievement.[10]

A study of CORE survey data use in Fresno, California, indicated that teachers did use the results for purposes such as identifying and addressing

problems with schools' discipline policies or corroborating (or not) teachers' perceptions regarding student engagement in learning.[11] At the same time, these educators recognized the surveys alone were insufficient for understanding the precise causes of student academic or behavior problems. Educators did not always understand what the surveys were measuring or what they should do in response to the data. Analyses of the CORE Districts' experience points to the value of these additional data sources but also makes clear that additional supports and guidance are needed to ensure their utility.

Four Questions to Guide the Design of Accountability Systems That Use Nonacademic Indicators

As the earlier examples make clear, nonacademic indicators currently play a minimal role, if any, in high-stakes accountability systems and tend to be used primarily for informing improvement efforts. However, some of the trends mentioned earlier could lead to a greater emphasis on nonacademic indicators for accountability purposes, so it is worth considering ways to ensure that efforts are informed by evidence and grounded in a careful exploration of the likely benefits and pitfalls. Accountability system design should be informed by standards for high-quality use of assessment data,[12] and they need to be attuned to contextual factors such as local priorities and the specific needs of the local population. In this section, rather than presenting general advice on accountability system design, which is available elsewhere,[13] I discuss four broad questions that those who develop and implement school accountability systems should ask.

What Is Your Purpose? Traditional "accountability" purposes might involve attaching rewards or sanctions to performance on indicators as a way of signaling priorities and motivating actors to change their behaviors. But, measures in accountability systems might also inform decisions about resource allocation, influence public support for the local school system, or identify students who need extra instruction. Some mandated consequences for poor performance under ESSA, for instance, are explicitly focused on identifying schools that could benefit from supports, even if they also create a perception of "shaming" those same schools.

Another potential purpose, intended or otherwise, is to inform school choice. Parents with access to high-quality information and extensive social networks already consider factors such as extracurricular offerings and safety when deciding where to send their children to school.[14] Any decision to incorporate a measure into an accountability system must be informed by its intended purpose, and evidence of validity must be gathered to justify that purpose.[15]

It is also important to predict how users might use data for unintended purposes and try to mitigate any potential harms from those unintended (and likely not validated) uses. A related consideration is the likely *consequences* of the measures, intended and unintended. Before adopting a new measure, ask what decisions this measure will inform and what consequences are likely to ensue and seek evidence related to those purposes and consequences. If such evidence does not exist, consider how to gather that evidence through pilots or detailed tracking of responses to the system.

The intended purpose should inform not only the choice of measures but also administration conditions (e.g., How important is test security?) and scoring considerations (e.g., Do you need a measure that produces subscores at the individual student level? Do you need individual-level scores at all?). Measures that are useful for informing instruction and other daily school functions are not necessarily the same ones that are suitable for accountability purposes.[16] Seemingly technical details can make a difference; for example, what kinds of scores are produced (norm-referenced, percentage proficient or above, etc.), and how are they weighted to produce an overall rating or index to inform decision-making. These details matter not just for the inferences the data will support but also for the messages the system sends to educators and others about what is valued. So do not avoid seemingly in-the-weeds questions about metrics and reporting when considering how to ensure the measures achieve their intended purposes.

Is Accountability the Only—or the Best—Way to Achieve Your Goals? In addition to clarifying the purpose up front, it can be helpful to consider

whether that purpose can be achieved through means other than adding an indicator to an accountability system. Schools' efforts to incorporate SEL into the curriculum, for instance, might be better served through the provision of instructional guidance, professional development, and supports for formative assessment practices that support teachers' instruction without the risks associated with high-stakes measurement. This question is especially relevant to ask about nonacademic constructs for which it is difficult to find a measure that has evidence of validity and reliability for the intended purpose and that is not easily corruptible in the face of high stakes.

System developers should be especially wary of the likelihood that an expanded system might result in excessive complexity or breadth that confuses rather than informs educators and other users of data from the system. Nonacademic measures might be used to label or sort students in ways that are inconsistent with their intended uses. Expansion is also likely to increase the costs of the system—not only the financial costs associated with developing and purchasing, administering, scoring, and reporting on measures but also the time the administration's use of these measures takes away from classroom instruction or principals' school leadership activities.

It is also worth considering the extent to which existing academic indicators draw on nonacademic processes or outcomes. A promising strategy for promoting SEL, for instance, is to integrate practices that promote social and emotional development into academic instruction.[17] This integration could be supported by assessments that incorporate skills such as collaboration or persistence, and new technology platforms are increasingly available to accommodate complex items. Other academic measures, such as GPA, provide evidence of students' attainment of a combination of academic and nonacademic (e.g., effort and persistence) skills.[18] The fuzziness of the academic and nonacademic boundaries means that even an accountability system that consists exclusively of academic indicators almost certainly captures a broader range of student competencies. Of course, it is impossible to disentangle the academic and nonacademic effects on an individual indicator so there might be good reasons to include standalone nonacademic indicators. But in the interest of simplicity and parsimony, it is wise to pause before expanding the accountability system and consider whether adding indicators is the most prudent strategy for achieving your goals.

Who Should Be Involved in Selecting Indicators? Changes to accountability systems can be contentious. In a recent AEI series on SEL, several authors provided advice for SEL advocates based on lessons from the Common Core experience and high-stakes testing more generally.[19] One common lesson was the need to engage with the right group of stakeholders throughout the decision-making process. Effective engagement can help groups feel confident that their views are represented in decision-making, even if they are not fully satisfied with the end result. It can also identify potential flash points early enough in the process to enable system developers to fix them rather than waiting until the system is launched and becomes a political minefield.

One of the most important reasons to seek broad and diverse input is that it can help promote goals related to equity, inclusion, and cultural responsiveness. Despite broad agreement on the value of nonacademic competencies and supports for all students, Robert Jagers and colleagues point out, "Questions have been raised about whether guiding frameworks, prominent programs, and associated assessments adequately reflect, cultivate, and leverage cultural assets and promote the well-being of youth of color and those from under-resourced backgrounds."[20] Broad stakeholder involvement could lead to key discussions around the appropriateness of indicators for all students. It could also result in more informed interpretation and use of data, such as through exploring ways to connect data from the indicators to information about structural barriers and opportunity gaps. Such connections could help prevent inaccurate inferences about the reasons for gaps and could point to potential solutions.

How Will You Ensure Data Are Useful? Merely collecting and reporting on data can change practice, especially by signaling priorities. However, gathering data that end up not being informative is a wasted opportunity. Before adopting new measures, make sure you have thought about *who* will use this

information and *how* they will go about doing so. The earlier point about clarifying the purpose is key, but merely describing the purpose of the measure does not guarantee that users will interpret the data and use them for decision-making in ways that promote the desired goal.

It will be crucial to ensure you have the plans, resources, and other supports to respond to information from the measures. For example, if you add measures of student well-being, what will you do if the results raise alarm bells for one or more children? Who will be responsible for acting on such information, and what resources will they have to respond appropriately? Questions about data use also have implications for decisions about the metrics you will develop (e.g., whether to provide a percentage above a cut score or a more continuous scale score) and the level of detail in the data provided to each stakeholder group. Trade-offs are inevitable; more fine-grained data can support more-informed decision-making but are also more likely to raise privacy concerns.

Do not assume teachers, principals, or district superintendents know how to respond effectively to nonacademic information. Even if they can understand what the data are telling them (e.g., one school in the district has a toxic climate or sixth-graders in the district show a large decline in self-efficacy), the specific steps they should take in response to these insights might not be evident. Existing published guidance on how to use measures such as SEL and climate surveys can provide a first step toward developing a plan for using data to inform practice.[21] Ideally, the group responsible for using accountability data should have access to information or technical assistance that provides research-informed guidance about next steps to take in response to these data.

Decision makers should be especially wary of using nonacademic indicator data to make decisions that have significant consequences for students or adults. Professional testing standards make it clear that such decisions should be based on multiple sources of data and must be informed by available evidence of validity for specific uses.

Looking Ahead

Growing interest in nonacademic accountability indicators reflects a widespread desire among educators, families, and policymakers for better information and more appropriate incentives. But even the most thoughtfully designed accountability system poses risks along with benefits. Perhaps the most important consideration is that although many nonacademic skills and processes can be measured, we have little to no valid evidence to support the use of these measures in accountability systems, and we have many reasons to fear the effects of unanticipated consequences. Research on test-based accountability offers reasons to be wary of attaching stakes to new measures, so developers must tread cautiously.

Meanwhile, there is a clear need for a renewed public discussion about the roles and responsibilities of the K–12 public school system. Long pressured to focus on preparation for college and careers, the system now finds itself at the center of debates regarding its duty to prepare students for civic life. Civic learning, broadly conceptualized, includes a mix of academic (e.g., understanding of government and history) and nonacademic (e.g., sense of social responsibility) competencies. Efforts are under way to develop short- and long-term measures of these competencies, and eventually, they could be valuable additions to a broad-based approach to monitoring school performance.[22]

In the meantime, rather than rushing to expand accountability systems, policymakers should carefully consider what they want to accomplish and add measures in a thoughtful and parsimonious way. The effort can start by asking the right questions.

About the Author

Laura S. Hamilton associate vice president at ETS. Previously, she served as a senior behavioral scientist and distinguished chair in learning and assessment at the RAND Corporation.

Notes

1. Lynn Olson and Craig Jerald, *The Big Test: The Future of Statewide Standardized Testing*, FutureEd, April 2020, https://www.future-ed.org/wp-content/uploads/2020/04/TheBigTest_Final-1.pdf.
2. Laura S. Hamilton, Julia H. Kaufman, and Lynn Hu, *Preparing Children and Youth for Civic Life in the Era of Truth Decay: Insights from the American Teacher Panel*, RAND Corporation, 2020, https://www.rand.org/pubs/research_reports/RRA112-6.html.
3. A full discussion of the issue of racial bias in standardized testing is beyond the scope of this report. For a discussion of conditions that contribute to inequities, see National Academies of Sciences, Engineering, and Medicine, *Monitoring Educational Equity*, 2019, https://www.nap.edu/catalog/25389/monitoring-educational-equity.
4. National Academies of Sciences, Engineering, and Medicine, *Monitoring Educational Equity*, 25.
5. Those categories are academic achievement, achievement growth, graduation rate, and progress in attaining English-language proficiency.
6. Council of Chief State School Officers, *Identifying a School Quality/Student Success Indicator for ESSA: Requirements and Considerations*, January 2017, https://www.nciea.org/sites/default/files/publications/CCSSO_SQSS_Brief.pdf.
7. Phyllis W. Jordan and Paige Marley, "How Did ESSA's "Non-Academic" Indicator Get So Academic?," FutureEd, March 9, 2018, https://www.future-ed.org/how-did-essas-non-academic-indicator-get-so-academic/.
8. Phyllis W. Jordan and Laura S. Hamilton, *Walking a Fine Line: School Climate Surveys in State ESSA Plans*, FutureEd, January 2020, https://www.future-ed.org/school-climate-surveys-in-state-essa-plans/.
9. Julie A. Marsh et al., *Enacting Social-Emotional Learning: Practices and Supports Employed in CORE Districts and Schools*, Policy Analysis for California Education, April 2018, https://www.edpolicyinca.org/sites/default/files/Report_SEL%20Practices.pdf.
10. Michelle Nayfack et al., *Building Systems Knowledge for Continuous Improvement: Early Lessons from the CORE Districts*, Policy Analysis for California Education, November 2017, https://www.edpolicyinca.org/sites/default/files/building%20system%20knowledge.pdf.
11. Thomas Toch and Raegen Miller, *CORE Lessons: Measuring the Social and Emotional Dimensions of Student Success*, FutureEd, February 2019, https://www.future-ed.org/wp-content/uploads/2019/02/FutureEd_Core_Report.pdf.
12. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, "Standards for Educational and Psychological Testing," https://www.apa.org/science/programs/testing/standards.
13. Readers should consult other sources for guidance and will need to grapple with decisions that are not addressed in the content that follows, such as whether to measure status or growth, whether to report subgroup performance, whether individual scores are necessary, and how to communicate clearly about expectations.
14. Alex Spurrier, Chad Aldeman, and Jennifer O'Neal Scheiss, "Refocusing the Priorities of Accountability," Bellwether Education Partners, 2020, https://bellwethereducation.org/sites/default/files/Bellwether_Accountability-RefocusingPriorities_Final.pdf.
15. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, "Standards for Educational and Psychological Testing."
16. Robin Lake and Lynn Olson, *Learning as We Go: Principles for Effective Assessment During the COVID-19 Pandemic*, Center on Reinventing Public Education, July 2020, https://www.crpe.org/publications/learning-we-go-principles-effective-assessment-during-covid-19-pandemic.
17. Aspen Institute, National Commission Social, Emotional, & Academic Development, *From a Nation at Risk to a Nation at Hope: Recommendations from the National Commission on Social, Emotional, and Academic Development*, 2018, http://nationathope.org/wp-content/uploads/2018_aspen_final-report_full_webversion.pdf.
18. Elaine M. Allensworth and Kallie Clark, "High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions About Consistency Across High Schools," *Educational Researcher* 49, no. 3 (January 2020): 198–211, https://journals.sagepub.com/doi/10.3102/0013189X20902110.
19. See Deven Carlson, "No Child Left Behind, National Ambitions, and Local Realities: Implications for Social and Emotional Learning," American Enterprise Institute, September 24, 2019, https://www.aei.org/research-products/report/no-child-left-behind-national-ambitions-and-local-realities-implications-for-social-and-emotional-learning/; Karen Nussle, "GPS for Social and Emotional Learning from One Who Traveled the Common Core Highway," American Enterprise Institute, November 18, 2019, https://www.aei.org/research-products/report/gps-for-social-and-emotional-learning-from-one-who-traveled-the-common-core-highway/; and Heather L. Schwartz and Laura S. Hamilton, "Get Smart About Social and Emotional Learning Measurement," American Enterprise Institute, August 21, 2019, https://www.aei.org/research-products/report/get-smart-about-social-and-emotional-learning-measurement/.
20. Robert J. Jagers, Deborah Rivas-Drake, and Teresa Borowski, "Equity & Social and Emotional Learning: A Cultural Analysis," Collaborative for Academic, Social, and Emotional Learning, November 2018, https://casel.org/wp-content/uploads/2020/04/equity-and-SEL-.pdf.

21. Jonathan Schweig, Laura S. Hamilton, and Garrett Baker, "School and Classroom Climate Measures: Considerations for Use by State and Local Education Leaders," RAND Corporation, https://www.rand.org/content/dam/rand/pubs/research_reports/RR4200/RR4259/RAND_RR4259.pdf; and Jeremy J. Taylor et al., *Choosing and Using SEL Competency Assessments: What Schools and Districts Need to Know*, RAND Corporation, November 2018, https://measuringsel.casel.org/pdf/Choosing-and-Using-SEL-Competency-Assessments_What-Schools-and-Districts-Need-to-Know.pdf.

22. For an overview of work on civic learning in K–12 public schools, see Hamilton, Kaufman, and Hu, *Preparing Children and Youth for Civic Life in the Era of Truth Decay*.