

Why Multiple Measures?

by Jack Schneider
College of the Holy Cross



mciea
MA Consortium for Innovative
Education Assessment

In all 50 states, students and schools are held accountable for their performance by systems inadequate to the task. One of the primary shortcomings of these systems is not a lack of data but rather a narrowness of data. The educational measurement and accountability systems established over the last two decades rely almost entirely on student achievement scores on standardized tests, often supplemented with graduation rates and a range of proxies for postsecondary and career readiness.¹

Insofar as the data revolution in education has focused attention on disparities across racial and income groups, it has had a positive effect. Yet current school quality measures, which rely largely on achievement tests in two subject areas—math and English—are highly incomplete. As a result, they are highly problematic.

The most obvious shortcoming of current measurement and accountability systems concerns what they exclude. Despite the present focus on test scores, schools have long served many purposes and advanced multiple aims.² Consider, for instance, the goals for public education identified by John Goodlad in his “Study of Schooling” project in the late 1970s³—a project in which researchers observed more than 1,000 classrooms and surveyed thousands of parents, students, teachers, and principals.

More recently, Richard Rothstein and Rebecca Jacobsen distilled the goals of schooling into eight major categories. Then, they asked representative samples of school board members, state legislators, school superintendents, and the American public to assign a relative importance to the goals, inviting them to allocate 100 “points” in whatever way they wished across the eight areas.

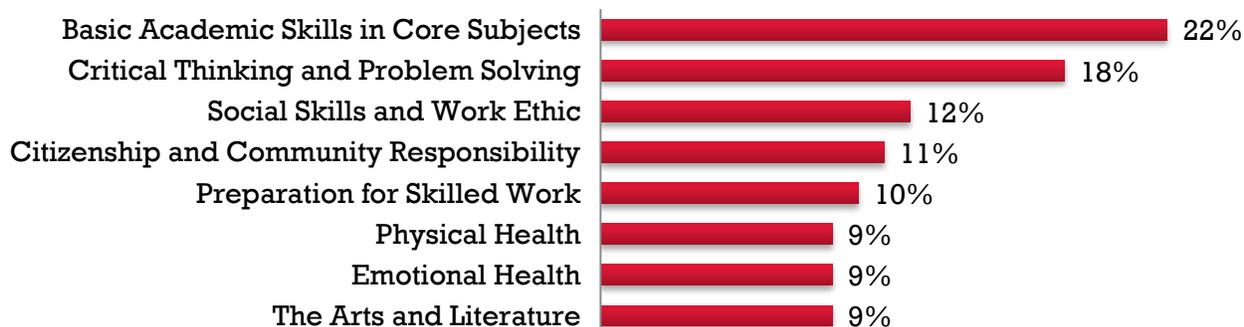
GOALS FOR PUBLIC EDUCATION

John Goodlad, “Study of Schooling”

- Mastery of Basic Skills and Fundamental Processes
- Intellectual Development
- Career Education
- Interpersonal Understandings
- Citizenship Participation
- Enculturation
- Moral and Ethical Character
- Emotional and Physical Well-being
- Creativity and Aesthetic Expression
- Self-Realization

As they found, responses were not only similar across groups, but also displayed balanced support for a broad range of aims.

Selected Americans' Views on Relative Importance of Public School Goals



Taken together, these perspectives (and similar perspectives expressed in a range of public polling research) may be interpreted as a broad consensus on the mission of public education. That is, a cross-section of stakeholders agree that schools have multi-pronged responsibilities for preparing academically knowledgeable, critically thinking, socially responsible, hard-working, physically healthy, and cultured citizens. That is a tall order, and it goes without saying that current measurement and accountability systems fail to address most of these expressed values. Additionally, as research suggests, various elements of school quality are not intrinsically aligned, indicating that a measurement system designed to capture only some elements of school quality will not necessarily capture others.⁴ At the most basic level, then, we have a problem: existing systems are not measuring all of what matters in public education, and they are holding schools accountable for only a narrow slice of their full mission.

More practically, in addition to their failure to reflect public attitudes about public schools, these constrained accountability systems also fail to deliver the full range of information relied upon by school leaders, policymakers, and the public. Current systems do little to meaningfully distinguish between successful and struggling schools, and they offer little in the way of actionable information about where assistance is needed. This is a powerful rationale for multiple measures.

Existing measurement and accountability systems are flawed not only in what they exclude, but also in what they *include*. As research has repeatedly documented, standardized test scores correlate quite strongly with student demography. Thus, while it is undoubtedly important to track academic

achievement, it is also the case that test scores often indicate more about a student's neighborhood and home life than about his or her educational experience.⁵ Relative to family background characteristics, school factors pale in comparison—explaining only about 20 percent of achievement.⁶ And though many measures of school quality will correlate with demography in some way, test scores appear to have the strongest correlation. In fact, none of the measures employed in an MCIEA pilot study correlated with student demography as strongly as test scores.

The relationship between test scores and demography distorts what existing data systems tell us about school quality. But the consequences of this correlation are intensified by the attachment of high-stakes accountability mechanisms to these flawed measurement systems—a coupling that has produced a number of troubling unintended consequences for public schools. For example, because schools are held accountable for a narrow set of scores—generally on math and reading tests in grades 3-8, as well as one year of high school—school leaders have responded rationally: by narrowing the curriculum. Arts, history, science, health, and other aspects of a diverse curriculum have been cut back dramatically over the past two decades, while emphasis on test-aligned math and English instruction has been ratcheted up. As a result, school has become less engaging and more stressful for students.⁷ And unfortunately for them, research indicates that increases in standardized test scores, though they may be associated with greater acquisition of content knowledge, may not be associated with cognitive growth.⁸

Perhaps the most serious unintended consequences associated with current measurement and accountability systems have to do with sanctions and stigmas. In most states, schools are responsible for bringing all subgroups of students—across race, gender, and class—to a level of defined “proficiency.” If they do not, the state intervenes by imposing sanctions, penalties, and eventually school shutdowns. But because low-income and minority students are likely to score lower on standardized tests, their schools are far more likely to be stigmatized by state intervention, or closed down, regardless of how much their scores have improved. Such stigmas and penalties create churn in school staff, as teachers flee or are fired. They send a message to students that they are on a dead-end track.⁹ And by further scaring away well-resourced and quality-conscious parents, such data intensify segregation. The young people most harmed by all of this are, it goes without saying, our least advantaged.

Multiple measures are not a perfect solution for the challenge of using data to support and strengthen schools. After all, no map can fully capture reality. But it does appear that we can ameliorate many of the most obvious flaws in present measurement and accountability systems by expanding the number of school quality indicators included.

Footnotes

¹ Marga Mikulecky and Kathy Christie, *Rating States, Grading Schools: What Parents and Experts Say States Should Consider to Make School Accountability Systems Meaningful*. Education Commission of the States (2014).

² David Figlio and Susannah Loeb, "School Accountability," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education* (The Netherlands: North-Holland, 2011): 383-417; Helen Ladd and Susannah Loeb, "The Challenges of Measuring School Quality," in Danielle Allen and Rob Reich, eds., *Education, Justice, and Democracy* (Chicago, IL: The University of Chicago Press, 2013): 19-42; Jack Schneider, *Beyond Test Scores: A Better Way to Measure School Quality* (Cambridge, MA: Harvard University Press, 2017).

³ See John Goodlad, *A Place Called School* (New York: McGraw-Hill, 1984).

⁴ Russell W. Rumberger and Gregory J. Palardy, "Test Scores, Dropout Rates, and Transfer Rates As Alternative Indicators of High School Performance," *American Educational Research Journal* 42, no. 1 (2005): 3-42.

⁵ Sean F. Reardon, "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations," *Whither Opportunity* (2011): 91-116; Pamela E. Davis-Kean, "The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment," *Journal of Family Psychology* 19, no. 2 (2005): 294.

⁶ Dan D. Goldhaber, Dominic J. Brewer, and Deborah J. Anderson, "A Three-Way Error Components Analysis of Educational Productivity," *Education Economics* 7, no. 3 (1999): 199-208; Barbara Nye, Spyros Konstantopoulos, and Larry V. Hedges, "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26, no. 3 (2004): 237-257; Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica* (2005): 417-458; Brian Rowan, Brian, Richard Correnti, and Robert Miller, "What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools," *Teachers College Record* 104, no. 8 (2002): 1525-1567.

⁷ Natasha K. Segool, John S. Carlson, Anisa N. Goforth, Nathan Von Der Embse, and Justin A. Barterian, "Heightened Test Anxiety among Young Children: Elementary School Students' Anxious Responses to High-Stakes Testing," *Psychology in the Schools* 50, no. 5 (2013): 489-499.

⁸ Amy S. Finn, Matthew A. Kraft, Martin R. West, Julia A. Leonard, Crystal E. Bish, Rebecca E. Martin, Margaret A. Sheridan, Christopher F.O. Gabrieli, and John D.E. Gabrieli, "Cognitive Skills, Student Achievement Tests, and Schools," *Psychological Science* 25, no. 3 (2014): 736-744.

⁹ See, for instance, David Scharfenberg, "Boston's Struggle with Income Segregation," *The Boston Globe*, March 6, 2016, <https://www.bostonglobe.com/metro/2016/03/05/segregation/NiQBy00TZsGqLnAT0tHsL/story.html>