



How (and Why) to Build a Better Measure of School Quality

Jack Schneider, University of Massachusetts-Amherst

For the past decade since the enactment of No Child Left Behind, policymakers and the public have had access to standardized test scores for all public schools. This allows schools to be evaluated and compared with more than anecdotes or comments passed by word of mouth. Still, changes in student test scores constitute a very narrow measure of what schools do and how well they do it. Overreliance on this one measure alone can misrepresent or even stigmatize otherwise effective schools. My colleague Anil Nathan and I have used a range of data collected in Massachusetts to develop a more comprehensive picture of school effectiveness. We aim to offer better tools to reformers working to improve schools and make them more accountable – and provide parents better information as they make enrollment decisions for their children.

Unintended Consequences from Narrow Measures

Rating schools only according to a narrow set of quantitative indicators leads to unfortunate consequences. As the social psychologist Donald Campbell wrote in 1976, “the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.” Using test scores to rate schools has produced such distortions. Educators have gamed such measures by narrowing the curriculum, teaching-to-the-test, and excluding low-performing students from testing – even, in some cases, by encouraging cheating.

Narrow measures also send an inaccurate message about which schools are succeeding and which are failing – exacerbating inequality in the process. As scholars have revealed, raw standardized test scores often reflect, primarily, the social and economic backgrounds of students. Yet parents may view such scores as indicators of how successful schools are in helping students do better. Consequently, even when schools whose students have low scores promote more *growth* for those students, the schools can be stigmatized as failures. And schools can be over-optimistically rated as high-performing when, in fact, their test scores merely reflect the privileged backgrounds of their students. Seeking the best schools for their children, quality-conscious parents with access to resources steer their offspring into schools with high test scores, and in the process create a self-fulfilling prophecy in which the schools with the best prepared students get more of them – which leads to more impressive test scores, even if the schools themselves are not making much difference. By contrast, schools with students who struggle on tests will continue to get relatively poor scores on this narrow measure.

Building a Better Measure

Is it possible to measure school quality in a more effective manner that avoids, or at least minimizes, these unintended consequences? There are two good ways to proceed:

- Use multiple measures of school quality. In performance management, the use of so-called “balanced scorecards” is a relatively standard practice. Applying this to education, a measure of school quality can include not only student standardized test scores, but also other indicators of features that parents, educators, and policymakers consider important in education – a measure of a school’s safety, its level of diversity, its academic climate, or even the availability of arts and music offerings.
- Allow users to place their own weights on school scorecard variables. For each measure of school quality, minimum thresholds can be set. But beyond that, various users of the scorecard can indicate the school qualities most important to them and see how schools rank with their weightings taken into account. Schools can still be compared, but without creating a single rank-ordered list of schools. This avoids improperly stigmatizing schools that do very well in some key ways – such as helping struggling students – and also assists parents and others trying to find schools that fit their values and meet the specific needs of children in varied circumstances.

In our research, we designed the “Dreamschool Finder” ranking tool that provides users with information about all public schools offering instruction in Kindergarten through twelfth grade in the state of

Massachusetts. Pulling from the available statewide data, we included more comprehensive measures than raw standardized test scores – giving users information about average student growth in mathematics, average student growth in English Language Arts, school climate, college readiness, school resources, and diversity.

Users of the Dreamschool Finder are able to weight the different measures of school effectiveness according to their own values – generating their own customized picture of school quality. This balanced scorecard approach reveals that various schools in Massachusetts excel on different measures. And the result is a more inclusive, and likely more accurate, portrait of effective schools.

Future Directions

Our project for Massachusetts had to rely on imperfect available data about schools, which we believe can be easily improved by state legislators and school administrators. Yet even despite our limitations, the Dreamschool Finder offered new perspectives on the varied ways in which schools can perform well – offering a more balanced picture of excellence than ranking schools on a single, narrow measure. Our approach also aligns better with what parents and the public want, because they, too, value a variety of forms of school excellence.

Our trial research on a more inclusive measure of school excellence puts us in a good position to suggest improved measures of school performance to policymakers, not just in Massachusetts but across the country. Public dialogue about improving schools – locally, at the state level, and nationally – needs to press ahead and become more nuanced and thoughtful. In order for that to happen, though, better evaluative measures of the variety of things schools do must be developed and used in districts, states, and in federal policymaking. No Child Left Behind's call for regular evaluation of school performance was a start, but to do such evaluation well, we must create and deploy a full array of meaningful measures.

Read more in Anil Nathan and Jack Schneider, "Rationale for School Quality Metrics and Construction of Ranking Scores," Boston Globe School Ranking Project, 2013.