

The “Third Way” of Education Reform?

By **Brian A. Jacob**, Kennedy School of Government

Education reformers agree on the need for high-quality teaching, but they disagree sharply on how to identify good and bad teachers. The 2002 *No Child Left Behind Act* ushered in a new era of standardized test-based accountability, with state statutes now rewarding and sanctioning school staff based on student performance and ten states directly linking teacher pay to test scores.¹ Many parents and educators, however, continue to view high-stakes testing as overly focused on narrow educational goals. Parents often want teachers to be role models and classroom managers, to develop strong relationships, foster creativity, encourage self esteem, or to teach a broad curriculum. Standardized exams, they argue, fail to capture these values and may detract from their pursuit. Both views have merit, but the battles between them often lead to gridlock.

Using a variety of data sources not normally available to researchers, we find that principals are an overlooked resource in these debates. Principals reliably predict which teachers will perform best at improving student test scores. Moreover, compared to standardized exams or traditional determinants of teacher pay, principals do a better job at assessing teachers’

abilities to meet other goals that are valued by parents. Empowering principals with greater authority could thus be an effective way to increase accountability while respecting a wider range of educational goals and the particular circumstances of each classroom and school.

How Do Principals Measure Up?

As front-line managers in schools, principals must distinguish high-quality from low-quality teaching. In research with Lars Lefgren, we confidentially surveyed elementary school principals in a mid-sized school district, questioning them about their teachers on a variety of performance measures.² We then compared how well these subjective assessment measures of teachers matched the subsequent test-score gains in individual classrooms, teacher salaries, and the classrooms that parents request for their children the following year.

Conventional measures of teacher capabilities are weak predictors of student performance on standardized exams. The most comprehensive way to assess teachers based on student test scores takes into consideration students’ prior performance and their demographic characteristics. After constructing such a “value-added”

Brian A. Jacob

Brian Jacob is Assistant Professor of Public Policy at the John F. Kennedy School of Government, a Faculty Research Fellow in the Program on Children and the Program on Education at the National Bureau of Economic Research, and a Faculty Associate with the Center for Basic Research in the Social Sciences. He has previously served as a policy analyst in the New York City’s Mayor’s Office and taught middle school in East Harlem.

This Policy Brief integrates research from Brian Jacob’s papers listing in the “Related Publications” section at the end of the brief.

A. Alfred Taubman Center for State and Local Government

The Taubman Center and its affiliated institutes and programs are the Kennedy School of Government’s focal point for activities that address urban policy, state and local governance and intergovernmental relations.

Taubman Center Policy Briefs are short overviews of new and notable research on these issues by scholars affiliated with the Center.

*A. Alfred Taubman Center for State and Local Government
John F. Kennedy School of Government,
79 JFK St., Cambridge, MA 02138*

Telephone: (617) 495-2199

Email: taubman@ksg.harvard.edu

www.ksg.harvard.edu/taubmancenter

index of teachers ability to improve student test scores, we examine the relationship between teacher salaries and these value-added measures. We find that teachers with characteristics that lead to higher compensation – experience and educational attainment – do no better than their colleagues at improving student test scores. First-year teachers and untenured teachers do perform slightly worse, but better-paid teachers overall are not notably better at improving test scores.

Principals’ own subjective assessments of teacher capabilities are better predictors of teacher value-added performance than the factors that determine pay levels. Principals appear particularly good at identifying teachers

Many parents and educators view high-stakes testing as overly focused on narrow educational goals.

at either end of the spectrum of performance – i.e., those who will produce the largest and smallest value-added achievement gains in their schools. However, they have far less ability to distinguish differences among teachers in the middle.³ Upon closer examination, we find two reasons that principals may not do as well with this group. In determining the ratings they give teachers, principals appear to place too much emphasis on the recent performance of their teachers and do not sufficiently account for the differences in initial student ability across classrooms. Perhaps not surprisingly, a teacher’s prior value-added score is an even better predictor of subsequent student performance than either the teacher’s education and experience, or the principal’s assessment of the teacher’s quality.

What Do Parents Value in Education?

Parent satisfaction is an alternative to test scores as an indicator of teacher quality. We gauge parent satisfaction by the number of parent requests for placement of their child with a specific teacher. Roughly 30 percent of parents made such requests according to principals’ records from 11 of the 13 schools in the district that provided this information. We find that principals’ assessments predict which teachers will be requested by parents in their schools better than *either* traditional factors in teacher compensation – education and experience – *or* our own teachers’ value-added assessment measures. However, the types of teachers preferred by parents vary considerably depending on the demographic characteristics of the student population in a particular school.

On average, parents strongly prefer teachers that principals describe in our survey as good at promoting student satisfaction. Parents place less value on a teacher’s ability to raise standardized math or reading achievement – as measured either by test scores or principal assessments. Parent requests correspond more closely to principals’ subjective assessments of teachers’ ability to raise student achievement than to the value-added measure, which indicates either that parents prefer the qualities captured in the principal assessments more than the value-added measures or that parents can more easily observe the qualities identified by the principal.

These average preferences, however, mask striking differences across family demographics. Parents in lower-income schools strongly prefer teachers who principal assessments and value-added indices indicate are good at improving test scores, as opposed to those teachers that principals describe as good at increasing student satisfaction.⁴ Interestingly, they appear to place little value on the softer teacher qualities recognized by principals. The results are reversed for families in higher-

income schools. These parents strongly prefer teachers who principals describe as good at promoting student satisfaction, and place no value on teachers described as good at promoting student achievement.

It thus seems that what parents want from school depends largely on family circumstances. Insofar as children from more advantaged households are more likely to read well by the second or third grade, for example, their parents will not prioritize basic phonics instruction. Parents whose children are still struggling with basic literacy will place greater value on instruction that emphasizes basic skills.

This has important implications for current school reform strategies, especially those that rely on mechanisms of parent choice

Parents in lower-income schools strongly prefer teachers who are good at improving test scores, as opposed to those teachers that principals describe as good at increasing student satisfaction. The results are reversed for families in higher-income schools.

in selecting schools. Our findings suggest that communities will react differently to accountability policies, such as those embodied in the federal accountability legislation *No Child Left Behind* (NCLB), depending on their demographic makeup. Specifically, since schools where fewer students are eligible for income-tested free-lunch programs prefer teachers that promote student satisfaction over testing gains, we predict that higher-income communities will express greater dissatisfaction with the achievement emphasis of NCLB.

Our findings may also have implications for school choice programs. On one hand, our analysis suggests that low-income families not only value student achievement, but also will be able to recognize which teachers will best promote these goals. This finding may alleviate concerns that disadvantaged students will not benefit from school-choice programs. On the other hand, systematic differences in parent values may also help explain why school-choice programs can lead to greater segregation between racial and income groups. Other studies have shown that parents tend to choose schools with racially and socially homogenous environments for their children. Our findings suggest that parents may not desire segregation *per se*, but instead view the socioeconomic composition of a school as a signal about whether others at the school share their interest in particular types of curriculum or pedagogy. Of course, because we study the choice of teachers within schools, rather than the choice of schools within a district or metropolitan area, these conclusions must be considered tentative.

Finally, our analysis suggests caution in interpreting school-choice programs in general. Low-income families, for instance, may choose to attend a school with a high proportion of other low-income families and low average test scores because they believe that the parents in these schools have preferences that, like their own, prioritize student achievement over other goals. Conversely, high-income parents may choose to attend schools with high test scores not because those schools engage in the basic skills and test preparation that is most helpful for increasing scores, but for completely opposite reasons – namely, because the preferences of families in those schools signal that teachers will engage in less basic skills instruction and offer instead a broad curriculum and activities that increase student engagement.

Empowering Principals

Our findings suggest that principals should have greater authority for school-wide staffing decisions, as they do at most charter schools. Currently, principals have some authority over the promotion of untenured teachers, but virtually no role in determining the promotion or compensation of tenured teachers. In fact, for a host of complicated social and political reasons, principals have been reluctant to exercise the authority they do have over untenured teachers. It is rare, for example, for an untenured teacher in a large school district to be dismissed for mere incompetence. One immediate way that superintendents could improve schools is by setting high standards for teaching excellence and then creating an environment in which principals are not only allowed, but actually expected, to dismiss ineffective untenured teachers

Principals’ subjective performance evaluations of teachers may thus offer a way out of the “testing/accountability” versus “professional autonomy” dilemma. Insofar as principals can observe inputs as well as outputs, they may be able to ensure that teachers increase student achievement through improvements

in pedagogy, classroom management, or curriculum. Principals can also evaluate teachers on the basis of a broader spectrum of educational outputs than test scores alone.

New powers for principals should also include safeguards. One reason that teachers formed unions in the first place was to defend against abuse and patronage by earlier generations of superintendents and principals. Our own results suggest, for instance, that principals may discriminate against male and untenured faculty.⁵ If teachers are to be held accountable to principals, then principals must be held accountable by other mechanisms.

Another question is whether empowering principals might lead to the kinds of unanticipated negative effects as those from high-stakes testing. Our analysis takes place in a context where principals were not explicitly evaluated by their superiors on the basis of their ability to identify effective teachers. It is possible that moving to a system where principals had more authority and responsibility for monitoring teacher effectiveness would enhance principals’ ability to identify various teacher characteristics. On the other hand, it is possible that principals would be less willing to

“Teaching to the Test” and Cheating

Test score gains are not synonymous with improved student learning. One reason many people view the scores as an imperfect indicator of educational success is that the impressive gains produced by high-stakes testing can also be artificially induced by “teaching to the test” and even cheating. My own research on the introduction of high-stakes testing in Chicago, for instance, indicates that some portion of the achievement gains in certain grades likely resulted from teachers instructing in test-specific skills, de-emphasizing subjects without high-stakes exams like science and social studies, and placing more students into special education where their scores often do not count.⁶ Moreover, policies that attach stricter consequences for low test scores prompt outright cheating by teachers and administrators that can be discerned by tell-tale patterns of test answers along with unusually large test score jumps. Using such methods, Steve Levitt and I found serious cheating in at least 4 to 5 percent of elementary classrooms in the Chicago Public Schools, a finding largely confirmed by subsequent retesting that launched official investigations.⁷

honestly assess teachers under such a system, perhaps because of social or political pressures. Just as high-stakes exams spur cheating and "teaching to the test" (see side bar), high-stakes principal assessments could lead to perverse outcomes.

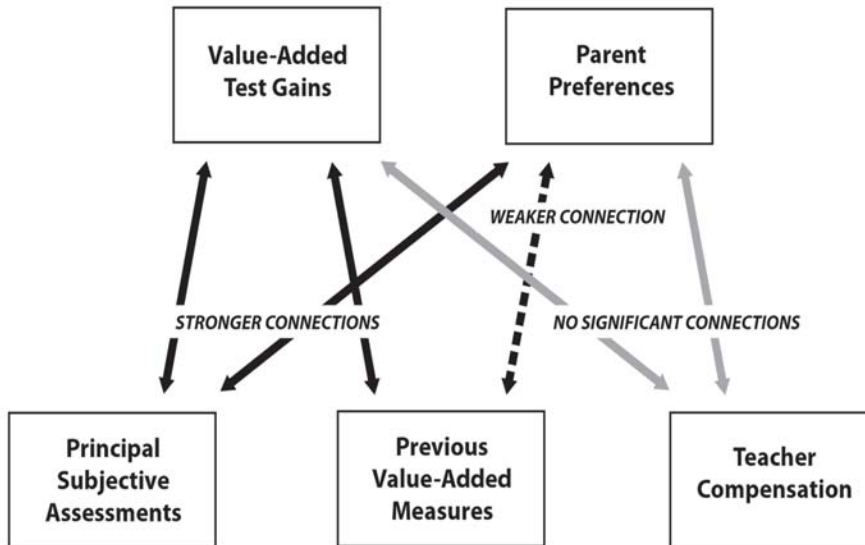
Nonetheless, our results suggest that switching from the current compensation system toward one that incorporates more

meaningful principal evaluations would improve both student achievement and parent satisfaction. The relative merits of principal-versus-test assessment is less clear.⁸ But empowered principals would likely enhance parent satisfaction, identify teachers who are particularly good or bad at promoting student achievement, and mitigate the negative consequences of high-stakes exams.

Figure 1: How Three Teacher Assessment Systems Predict Student Test Gains and Parent Preferences

The "value-added" measure of effectiveness is based on how much improvement students make on standardized tests from one year to the next. We adjust the scores to account for the fact that different teachers serve different types of students who may be more or less prepared for academic work. Thus, teachers are measured as producing more "added value" when they produce large testing gains for students that generally perform less well on these exams.

Principals record **parents requests** to have their child assigned to a particular teacher for the following school year. Roughly 30 percent of parents request a teacher each year.



Principals were asked to confidentially rate their teachers based on their ability to produce test-score gains, obtain student satisfaction, and other skills. Principal estimates of teachers' ability to increase student achievement predicted future student performance almost as well as value-added measures based on prior student achievement in a teacher's class. Parents in lower-income schools appear to value student achievement more than student satisfaction, while parents in higher-income schools seem to prefer teachers with high student satisfaction ratings.

Teacher compensation is traditionally determined through collectively bargained salary schedules that focus on years of prior experience and teacher educational attainment. Our findings show that, while small effects can be identified between individual education or experience factors, teacher compensation overall bears no relationship to either value-added test gains or parent preferences.

Endnotes

- 1 “Teacher Merit Pay Tied to Educational Gains,” *The New York Times*, October 4, 2005
- 2 The district, located in the western United States, gave permission for these February 2003 interviews and has requested to remain anonymous. The district is not unrepresentative of American public schools as a whole. While students in the district are predominantly white (73 percent), Latinos comprise 21 percent of the elementary population. Nearly half of all students (48 percent) in the district receive free or reduced price lunch. Testing achievement levels in the district are almost exactly at the average for the nation (49th percentile on the Stanford Achievement Test). The sample consists of 202 teachers in grades two through six.
- 3 Principals appear somewhat better at identifying top and bottom-quality teachers for math than reading.
- 4 Disadvantaged parents are relatively less likely to request any teacher. But their relative proclivity to choose teachers based on test-score results is particularly striking given that these parents are arguably less likely to be adept at inferring judgments based on comparative numerical data.
- 5 Principals do not report worse personal relationships between themselves and male or untenured teachers, but they give systematically give lower ratings to these teachers than the teachers’ value-added scores would otherwise dictate.
- 6 Once high-stakes consequences were introduced, students seem to have left fewer exam answers blank where there was no penalty for guessing. Fewer students also seem to give up in the final sections of exams.
- 7 We look for cases where particular teachers’ students experience unusually large test score gains, followed by unusually small gains or even declines in the following year. We also look for patterns of suspicious answer strings, such as identical blocks of answers for many students in a classroom, or cases where students are unable to answer easy questions correctly, but do exceptionally well on the most difficult questions. Following our findings, the Chicago Public Schools conducted extensive retesting of suspicious classrooms and controls groups. The results largely confirmed our findings and led to investigation of 29 classrooms.
8. Widespread implementation of value-added assessment of test scores might itself face obstacles because the approach implicitly designates certain groups as harder to teach at the time. It does this by statistically anticipating how students of certain races, gender, and income will have greater or lesser difficulty achieving testing gains.

RELATED PUBLICATIONS

“What do Parents Value in Education: An Empirical Investigation of Parents Revealed Preferences for Teachers,”

By Brian A. Jacob and Lars Lefgren, National Bureau of Economic Research Working Paper 11494 (2005) [<http://ksghome.harvard.edu/~bjacob/parents27.pdf>]

“Principals as Agents: Subjective Performance Measurement in Education”

By Brian A. Jacob and Lars Lefgren, National Bureau of Economic Research Working Paper (2005) [<http://ksghome.harvard.edu/~bjacob/principals35.pdf>]

“Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in Chicago Public Schools,”

By Brian A. Jacob, *Journal of Public Economics*, 89, 5-6 (2005) [<http://ksghome.harvard.edu/~bjacob/highstakes16.pdf>]

“Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating,”

By Brian A. Jacob and Steven D. Levitt, *Quarterly Journal of Economics*, 118, 3 (2003) [<http://ksghome.harvard.edu/~bjacob/cheating.pdf>]

“Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory”

By Brian A. Jacob and Steven D. Levitt, Brookings-Wharton Papers on Urban Affairs, William G. Gale and Janet Rothenberg Pack, eds. (Brookings Institution Press, 2003) [http://ksghome.harvard.edu/~bjacob/bw03cht_abs.htm]