

The Test of Seriousness

Student Achievement and Performance-Based Compensation

Due to a surge of retirements at a time of growing enrollment, the nation must increase teacher recruitment by *35 percent* over the next decade.¹ That is not just a challenge, but also an historic opportunity to attract more of America’s best and brightest into education. To accomplish that, we will need to begin to treat teachers like other professionals and offer them the opportunity to be rewarded for performance on the job regardless of their seniority.² Performance-based pay can also help better focus our schools on what matters most—student learning.

However, as policymakers consider performance-based pay plans, concerns about testing often arise. Some argue that it is simply unfair to evaluate teachers based on their students’ test scores. Others worry that tests cannot measure everything that matters, and some even suggest that tests measure *nothing* of value. Still others fear that incentives based on test scores will encourage teachers to react in negative ways, such as “teaching to the test” or “narrowing the curriculum.”

But student achievement is far too important to leave out of any proposal to value teachers by rewarding their performance in the classroom:

- While tests cannot measure every single thing that everyone might care about, they offer the best way to measure something we should *all* care about—whether students are becoming more knowledgeable and more skilled.
- Experts have devised fair and accurate ways to use student test scores to evaluate teaching.
- Despite alarmist rhetoric about testing, fears about unintended consequences are often misinformed or vastly inflated.

Test Scores Can Be Used Fairly

Many who argue that student assessments are an unfair way to evaluate teacher performance assume that schools would just use year-end test scores. “Say that at the *beginning* of the year, Mrs. Anderson’s students start out knowing a lot more than Mr. Franklin’s students, simply because they had better teachers in all the years before. What if Mr. Franklin works hard and his students make more progress, but they still end up scoring lower than Mrs. Anderson’s students on the tests in May?”

But that is not how modern performance-based compensation programs actually work. Instead, they measure *growth* in student achievement by taking into account a student’s skills and knowledge

upon entering a teacher's classroom and then look at how much the student gains by the end of the year. Growth measures provide a fair picture of each individual teacher's contribution to student learning.

National polls show that when teachers are asked about performance incentives based on student achievement *gains* rather than simple year-end test scores, their support rises considerably. In a 2003 poll by the research firm Public Agenda, 50 percent of beginning teachers rated such a proposal as "excellent" or "good," while only 15 percent rated it "poor."³

Some value-added methods go beyond that, taking advantage of an even broader range of information about students. A method pioneered in Tennessee and now widely used across the country examines at least three years worth of test score gains for each student. That is, it looks at each student's learning trajectory—how much a student grew over the two or more years prior to entering a teacher's classroom—before looking at how much the student's skills grow under that teacher's watch. The method predicts how much a student should gain each year based on his or her learning history, and sees how much she actually did gain.

The resulting "effectiveness scores" have proven a very reliable measure of teacher performance in that they are very stable from year to year. So there is no reason to be concerned that teachers who have a particularly strong or weak group of students in a given year will be disadvantaged under a performance-based compensation plan.

Many programs take additional precautions. For example, the Teacher Advancement Program (TAP), which uses the Tennessee value-added method, requires that a student attend a minimum number of days in a teacher's classroom for his or her scores to count in the analysis. TAP requires teachers to vote before a school can participate in the program, and it is now being implemented in nearly 150 schools across the country.

It might even be argued that value-added test-score gains provide the *fairest* way to know whether teachers are accomplishing what matters most in the classroom. Consider the following story told to the Education Trust by an administrator in a district that has tools to analyze student learning gains teacher by teacher:

The principal said, '[Mrs. Jones is] very negative. She's always complaining to me about her students, how they don't perform well enough, how they never live up to her expectations.' Well, after the meeting I went back and looked up the value-added data for that school. It turns out that Mrs. Jones was consistently the most effective teacher in the entire building. She was 'negative' only in the sense that she had very high expectations for her students and so was never really satisfied—and the students benefited tremendously as a result."⁴

Test Scores Are Meaningful

Tests and test scores have real life consequences for students, and assessments results provide valuable information about teachers.

Some critics claim that standardized assessments do not measure anything that really matters for students. But ***tests are increasingly a part of life beyond high school***, and not only for those who go on to higher education. For example, trade unions now rely on standardized tests to screen candidates for apprenticeship programs:

- The Sprinkler Fitters Union Local 709 in Los Angeles recently hired a testing company to create a math assessment as a way to screen applicants. “We don’t know if these kids can do math,” the union’s director of training said. “Even though they all have diplomas, some know algebra, while others have never even worked with fractions. In construction, math is a big part of everything we do. You have to be able to convert fractions to decimals, read plans, cut materials. It’s important that candidates have the skills to do that.”⁵
- The International Brotherhood of Electrical Workers (IBEW) and the National Electrical Contractors Association has created a standardized assessment that local apprenticeship programs can use to screen applicants. The assessment includes algebra problems as well as questions to measure reading comprehension.⁶

Question on Entrance Assessment for Electrical Apprenticeship

2. Consider the following formula: $y = 3(x + 5)(x - 2)$
Which of the following formulas is equivalent to this one?

A. $y = 3x^2 + 9x - 30$
 B. $y = x^2 + 3x - 10$
 C. $y = 3x^2 + 3x - 10$
 D. $y = 3x^2 + 3x - 30$

According to Jonathan Mitchell, the training director of the IBEW Local 490 electrical union in Concord, New Hampshire, last year half of the applicants failed the aptitude test. “It’s the difference between being in the ditch and being at the print table,” he says. “Last year we brought in two who just didn’t have the math, even after hours of working with them and showing examples. It makes you wonder, ‘How did they get that grade? How did they pass their high school math classes?’”⁷

Students who go to college are often surprised to find they have to take a standardized placement test *after* they get in. Many are even more surprised when they fail the test and have to waste valuable time and money taking remedial courses. Because more than half of entering freshmen enter California State University campuses unprepared, the system has begun offering an Early Assessment Program (EAP) for high school juniors to gauge whether they are on track to be ready for college. Last year 134,000 high school juniors *voluntarily* took the EAP mathematics test and 158,000 completed the English test. The majority of them failed.⁸

Moreover, education researchers and economists have documented that ***K-12 test scores predict the future success of students.*** For example, a trio of Johns Hopkins researchers recently noted, “achievement scores at any level of schooling predict success at the next level. This holds for high school completion, college attendance, college completion, and later successes in the labor market.”⁹

In fact, according to economist Eric Hanushek, “There is mounting evidence that quality measured by test scores is directly related to individual earnings, productivity, and economic growth. A variety of researchers documents that the earnings advantages to higher achievement on standardized tests are quite substantial.” He points to several recent studies that, taken together, suggest boosting student achievement scores by a standard deviation increases later earnings by 12 percent.¹⁰

Another study published last year found that students who made substantial test score gains in mathematics during high school had higher earnings seven years later than otherwise similar students who did not. “The high correlation between test scores and socioeconomic status suggests that one way to improve the skills and productivity of those at the bottom of the socioeconomic ladder is to improve their test scores.”¹¹

The nation stands to gain a lot from higher test scores, too. Hanushek estimates that if America could raise the skills of our students to just the middle of the pack of European nations over the next decade, our Gross Domestic Product would grow by two percent extra over 20 years and five percent extra over 30 years. That would mean an extra \$1.5 trillion in 2037 alone—more than triple what we currently spend on K-12 public education.¹²

If helping students make gains in their test scores will enable them to achieve more as adults, not to mention help America achieve more as a nation, shouldn't we value and reward teachers who do a good job of it? Any plan for performance-based compensation that takes no account of student achievement would be like judging an architectural blueprint without considering structural integrity. Other factors might matter, too, but at the very least an architect should ensure that every building has a strong foundation.

Finally, ***student test scores can provide a great deal of critical information about teachers.*** Over the past fifteen years, researchers studying teacher effectiveness based on value-added student test score gains have found:

- Teachers have a greater impact on student learning than any other school-related factor.¹³
- There are huge differences among teachers in their ability to produce student learning gains, even among teachers within the same school.¹⁴
- Students in low-income and high-minority schools are much more likely to get ineffective teachers and less likely to get effective ones.¹⁵
- If we provided low-income students with good teachers rather than merely average teachers for four to five years in a row, we could entirely close the achievement gap.¹⁶
- Giving students from any background several ineffective teachers in a row deals them a crushing blow from which they seldom recover.¹⁷

Savvy educators are beginning to find meaning in student achievement gains, too. Teachers at Boston's McCormick Middle School analyzed test results by classroom to determine whether some teachers were doing a better job helping students master particular mathematics topics. According to Harvard University researchers who studied the school, “The students of one teacher excelled on questions dealing with graphs; for another teacher, it was fractions; for a third, pre-algebra topics.” In response, the school's math teachers decided to carve out time to teach each other their effective methods.¹⁸

Given the valuable information student assessment results can provide, how can we ignore what they tell us about individual teachers working in our nation's classrooms?

Of course, some critics object that since tests are not given in certain grades and subjects (such as art), it would be unfair to use test scores to measure teacher effectiveness because not all teachers could be evaluated on that basis. But the answer is to strive to develop more tests and testing methodologies that can cover a broader range of teachers, and use alternative assessments until they are available. Where teachers *can* be evaluated using their students'

learning gains we should take the opportunity to do so. It would be foolish to discard valuable information simply because we do not have it for everyone.

Moreover, many programs use additional measures that can be used for all teachers, including those who do not work in grades or subject areas where there are annual student assessments. In the TAP program, half of the performance rewards are based on value added to student achievement and half are based on observations of teaching skills that take place 4-6 times per year.

Fears about Negative Consequences Are Inflated

Perhaps no criticism carries more weight than the claim that using test scores would cause an avalanche of terrible teaching practices in America's schools. But hard facts paint a much different picture than the heated rhetoric conjures up.

“Teaching to the Test”

The words “teaching to the test” are widely used but seldom clearly defined, causing much confusion. Indeed, last year the *Washington Post* writer Jay Mathews argued that if teaching to the test simply means aligning classroom instruction and curriculum with rigorous education standards, it is a good practice that should be supported rather than criticized.¹⁹

Stuart Yeh studied the impact of high-stakes testing in Minnesota and found that teachers thought testing *improved* the quality of their instruction. According to a summary by the Center for Public Education:

Eighth-grade teachers reported rewriting the curriculum to align with the state test because the eight strands of math it assesses are, as one teacher said, “Necessary skills for kids.” A principal told Yeh that “some of those fluffy extraneous things” were eliminated from the curriculum. Math and science teachers embraced the opportunity to teach reading in their subject areas. One said, “Drilling is very little help, because I see the math test is essentially a reading test.” In general, by a two-to-one margin, the Minnesota teachers thought that the impact of state testing was positive.²⁰

Some might still worry that teachers would spend most of their time drilling students on questions likely to be found on the tests. But it turns out there *actually is no true incentive* for teachers to do so.

Researchers at the University of Chicago looked at nearly 2,000 classroom assignments to evaluate whether they called for “authentic intellectual work” from students—such as applying basic skills and knowledge to solve new problems; describing ideas and solutions in ways that require strong communication skills; and producing work related to the real world beyond the classroom—above and beyond basic “skill and drill” answers.

In classrooms where teachers gave more assignments that called for high-level intellectual work, students logged gains on a nationally norm-referenced test that *exceeded* the national average by 20 percent. At the same time, students who were given mostly simple drills and few authentic assignments gained much less than the national average. The same pattern emerged when they examined student gains on a second statewide assessment. The researchers concluded:

Fears that students will score lower on conventional tests due to teacher demands for more authentic intellectual work appear unwarranted. To the contrary, the evidence indicates that assignments calling for more authentic intellectual work actually improve student scores on [standardized] tests.²¹

The choice between good instruction and good test scores is a false one. It is in teachers' interests as well as their students' to aim high and teach well.

“Narrowing the Curriculum”

So far, fears about drastic cuts to science and social studies have turned out to be misplaced, too. A national survey of teachers found a small decline in instructional time devoted to science, social studies, and math, along with an increase in English, from 1999-2004, part of a general trend that has been occurring since the 1980s.²² Even so, in March the U.S. Department of Education reported significant *gains* in history achievement on the National Assessment of Educational Progress.²³ Some experts speculate that because reading is central to learning other subjects, better reading skills have translated into more learning in those subjects.

Also, we know that highly effective teachers integrate reading, writing, and math into their curricula of non-core subject areas. Instead of devaluing non-core subjects, school leaders can focus on cross-content integration in those subjects, thus broadening the curriculum. Integration allows for students to more fully demonstrate what they know and are able to do in non-core subject areas, while developing the cross-content skills necessary to perform on standardized tests.

Finally, research shows that schools would get *lower* gains on reading tests if they narrowed the curriculum too much. Cognitive scientists have found that studying a wide range of subjects helps students build a broad vocabulary and extensive “background knowledge,” both of which contribute to better reading comprehension skills and, in turn, better reading scores.²⁴ That means schools have a *disincentive* to cut non-tested subjects and a positive incentive to teach a balanced curriculum, since that will help students develop better reading skills and make bigger gains on reading tests in the long run.

In short, the bad practices we fear—“teaching to the test” and “narrowing the curriculum”—actually *disadvantage* teachers who resort to them since they ultimately yield *less* growth on standardized tests. No doubt some teachers feel pressure to take shortcuts. But the answer is to educate and inform teachers so they understand that good instruction—the kind of classroom teaching we all want for students—is the key to success on all measures that matter.

Valuing Teachers for What Matters Most

No one is arguing that tests are an absolutely perfect way to measure everything we might care about. Nor is value-added an absolutely perfect way to measure teacher effectiveness. But a focus on results can certainly help improve the system most districts use now, which reveals almost nothing about true effectiveness and instead compensates teachers based merely on experience and advanced degrees. One recent study found that teachers who earn master's degrees are, on average, *less effective* at improving student achievement than those who do not.²⁵

An important goal of K-12 education is to increase student learning, so it seems logical that an important measurement of teacher effectiveness is how much their students learn—not how high students score, but by how much they improve.

It is time to identify teachers who are truly effective at improving student's skills, provide them opportunities to help their peers, give them incentives to take on tough assignments, and value them enough to reward their results in the classroom. **Today, most school districts in America do almost none of those things.** And that will never change if leaders fail to capitalize on the critical information that student assessment results can provide.

Endnotes

¹ Gordon, R, Kane, T., & Staiger, D. (2006). *Identifying Effective Teachers Using Performance on the Job*. Washington, DC: The Brookings Institution. (p. 12)

² Hoxby, C. M. & Leigh, A. (2004, May). Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States. *American Economic Review*, 94(2), 236-240.

³ Farkas, S., Johnson, J., & Duffett, A. (2003). *Stand by Me: What Teachers Really Think about Unions, Merit Pay and Other Professional Matters*. New York: Public Agenda. (p. 53)

⁴ Carey, K. (2004). *The Real Value of Teachers: If Good Teachers Matter, Why Don't We Act Like It?* Washington, DC: The Education Trust. (p. 8)

⁵ See <http://www.wonderlic.com/resources/Results/sprinklerFittersUnion709.asp>.

⁶ See <http://www.njatc.org/training/apprenticeship/index.aspx>.

⁷ Personal interview.

⁸ See <http://www.calstate.edu/EAP/>

⁹ Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007, April). Lasting consequences of the summer learning gap. *American Sociological Review*, 72(2), 167-180. (p. 168)

¹⁰ Hanushek, E. A. (2006, August). Alternative school policies and the benefits of general cognitive skills. *Economics of Education Review*, 25(4), 447-462.

¹¹ Rose, H. (2006, August). Do gains in test scores explain labor market outcomes? *Economics of Education Review*, 25(4), 430-446. (p. 445)

¹² Hanushek, E. A. (2006, August). Alternative school policies and the benefits of general cognitive skills. *Economics of Education Review*, 25(4), 447-462. (p. 454)

¹³ Sanders, W. L. & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.

¹⁴ Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004, Fall). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237-257.

¹⁵ Gordon, R, Kane, T., & Staiger, D. (2006). *Identifying Effective Teachers Using Performance on the Job*. Washington, DC: The Brookings Institution.

¹⁶ Hanushek, E. (2006). Alternative school policies and the benefits of general cognitive skills. *Economics of Education Review*, 25, 447-462. (p. 456)

¹⁷ Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

¹⁸ Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using student-assessment results to improve instruction: Lessons from a workshop. *Journal of Education for Students Placed At Risk*, 10(3), 269–280. (pp. 269–270)

¹⁹ Mathews, J. (2006, February 20). Let's teach to the test. *The Washington Post*, p. A21. Retrieved June 30, 2006, from <http://www.washingtonpost.com/wp-dyn/content/article/2006/02/19/AR2006021900976.html>

²⁰ Center for Public Education, "Research review: Effects of high-stakes testing on instruction." Retrieved July 16, 2007, from

[http://www.centerforpubliceducation.org/site/c.kjJXJ5MPIwE/b.1536671/k.9B6A/Research review Effects of highstakes testing on instruction.htm#q6a](http://www.centerforpubliceducation.org/site/c.kjJXJ5MPIwE/b.1536671/k.9B6A/Research%20review%20Effects%20of%20highstakes%20testing%20on%20instruction.htm#q6a)

²¹ Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence?* Chicago: Consortium on Chicago School Research. Retrieved June 30, 2006, from <http://www.consortium-chicago.org/publications/pdfs/p0a02.pdf>

²² Morton, B. A. & Dalton, B. (2007, May). *Changes in Instructional Hours in Four Subjects by Public School Teachers of Grades 1 Through 4*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

²³ National Center for Educational Progress, National Assessment of Educational Progress 2006, http://nationsreportcard.gov/ushistory_2006/.

²⁴ Jerald, C. (2006, August). *The Hidden Costs of Curriculum Narrowing*. Washington, DC: Center for Comprehensive School Reform and Improvement.

²⁵ Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007, January). *How and Why Do Credentials Matter for Student Achievement?* Cambridge, MA: National Bureau for Economic Research. (pp. 32-33)