

THE CENTURY FOUNDATION

Issue Brief

EIGHT REASONS NOT TO TIE TEACHER PAY TO STANDARDIZED TEST RESULTS

BY GORDON MACINNES

INTRODUCTION

Education Secretary Arne Duncan insists that states should be ready to mandate that teacher evaluation and compensation be pegged to how well their students perform on standardized tests. In the proposed rules for the Race to the Top Fund—the federal program that is seeking to distribute \$4.3 billion in aid to states that are implementing innovative and ambitious plans for increasing student achievement—having a data system that does not block teacher and student identifications is one of just three absolute conditions for receiving funding. The Department of Education will not even read a state’s application if it stumbles on this mandate.

Here are eight reasons why everyone’s time, energy, and tax dollars would be better spent elsewhere.

Reason #1: Tying test scores to teacher compensation suggests that teachers are holding back on using their experience, expertise, and time because they are not being paid for the extra effort.

Instead, the evidence is strong that most teachers simply do not know what to do when confronted with concentrations of poor children who are unprepared for the grade level or content taught. The surge in students from non-English speaking families further complicates teachers’ jobs. In the absence of helpful guidance from those who supervise and evaluate them, teachers fall back on what they have done all along, even if it is not particularly effective.

What is more worrying, and should be the object of reform efforts in Washington and state capitals, is that the leaders of most urban districts and schools—those who set the tone and the boundaries of practice—do not know what to do to improve the educational prospects of poor children. Moreover, there is no evidence that the policymakers on boards of education, in legislatures, or in departments of education have a clear vision of how to educate concentrations of poor children effectively. Otherwise, after forty years of serial reforms, the gap between poor and affluent students would have narrowed more than it has.

The Century Foundation conducts public policy research and analyses of economic, social, and foreign policy issues, including inequality, retirement security, election reform, media studies, homeland security, and international affairs. The foundation produces books, reports, and other publications, convenes task forces and working groups, and operates eight informational Web sites. With offices in New York City and Washington, D.C., The Century Foundation is nonprofit and nonpartisan and was founded in 1919 by Edward A. Filene.

HEADQUARTERS: 41 EAST 70TH STREET – NEW YORK, NY 10021 – 212.535.4441 – 212.535.7534 (FAX) – INFO@TCF.ORG – WWW.TCF.ORG
D.C. OFFICE: 1333 H STREET, NW – 10TH FLOOR – WASHINGTON, D.C. 20005 – 202.387.0400 – 202.483.9430 (FAX) – INFO@TCF.ORG – WWW.TCF.ORG

Reason # 2: The standardized tests in most states are lousy and so are the standards they are designed to measure.

Many well-organized groups have raised legitimate concerns that standardized testing can be detrimental to students by undermining the richness and breadth of their education, and drawing a false picture of academic achievement. But it is Secretary Duncan who has said that he believes that existing tests are weak and easily manipulated by many states, and that state academic standards are too plentiful, too vague, and too easy.

One of the Obama administration's four top objectives for spending stimulus funds is to improve the rigor of academic standards and the reliability of the standardized tests. Secretary Duncan is prodding states to join the voluntary effort managed by the National Governors Association and the Council of Chief State School Officers to come up with common core standards and to collaborate by developing a single, reliable set of tests.

But since this voluntary effort will take several years to work its way through fifty states, lousy standardized tests should not be used now to evaluate and compensate teachers.

Reason #3: The idea of compensating teachers individually in order to differentiate their performance from their school colleagues defeats a principal tenet of good instruction—that teachers need to learn from one another to solve difficult pedagogical challenges.

The Department of Education is praising higher-performing charter schools such as KIPP that strongly emphasize a culture of teacher cooperation. Establishing such a culture is one of the options under the department's proposed rules that states and districts can use to turn around failed schools. The evidence from especially effective districts such as Montgomery County, Maryland, and Union City, New Jersey, is powerful that frequent teacher collaboration is essential to improved teaching and learning.

Good instructional practice can be thought of as being much more like professional basketball than professional golf. The legendary Boston Celtics won thirteen championships through their unselfish play and relentless team defense. The benchwarmers, who sharpened the game of the starters, received the same championship rings as Bill Russell and Bob Cousy. Golfers, in contrast, are paid only when they score better than other competitors.

The last thing we need is to isolate city teachers from each other by introducing test-score driven competition with their colleagues.

Reason #4: Most teachers do not teach a grade or subject that is subject to standardized testing.

By itself, this is reason enough to be wary of the Department of Education's proposal. No framework or advice, or even a vague notion, is offered by the department as to how a kindergarten or French teacher would be evaluated as a part of the new scheme. If the idea is that "we're-all-in-the-same-boat" will govern the evaluation of the entire school, then the teacher-student identification capability need not be one of three absolute conditions for application.

Reason # 5: Even reliable standardized tests are valid only when they are used for their intended purposes.

The state tests mandated by NCLB for grades 3–8 and one year of high school are to measure how well *students* have mastered their state standards. They are not designed to measure how well teachers teach.

By comparison, to get a driver’s license, one must typically pass a vision and written test of laws and regulations, as well as a driving test with an examiner—who would favor granting the license based on the eye exam alone? The SAT was designed to help predict how well individual high school seniors would perform academically in their first year of college, not to judge the quality of their high schools. Otherwise, one might conclude that Mississippi’s high schools are far superior to those in Massachusetts, since their students’ mean SAT scores were 53 points higher in reading and 28 points higher in math. Accurate information used inappropriately can mislead everyone and lead to ill-conceived actions.

Reason #6: A key assumption of using test scores to judge teachers is that students are randomly assigned, first, to schools, and, second, to classes. Neither is true.

This is a methodological problem that has bedeviled the evaluation of charter and magnet schools, since it so difficult to assign a numerical value for having parents who seek better schools for their children. Ask principals of affluent suburban schools if they could get away with random classroom assignments. The situation is the same in city schools. If Mr. Jones knows a little Spanish, then he should take the recent immigrant from Mexico who speaks no English; the recently hired Ms. Cuccio can take the below-grade level readers; while the veteran complainer Mrs. Green gets most of the kids already reading at grade level. This “selection bias” contaminates the value and validity of much statistically driven research.

Consider the effect that high student mobility might have on test scores as well. My own analysis of the bottom 5 percent of New Jersey elementary schools found that student mobility ranged from the high teens to 50 percent (versus a state average of 11 percent). In such a system, there is no way to measure what each teacher may give up or inherit with departing and arriving students. No one has determined satisfactorily how to measure the effect of, say, non-English speaking immigrants taking the place of native English-speakers in November on either the student’s academic performance or the teacher’s effectiveness.

Reason #7: State data systems are in their infancy. It turns out that it is harder, is more expensive, and takes longer for states to produce reliable, accurate, and secure longitudinal data on students and teachers than widely assumed.

Better evaluation is built on tracking the progress of every student over time. With powerful data systems, it is now possible to incorporate much relevant information beyond race, economic status, and scores on state tests, such as who taught each student each year or in each course. Sensibly, Secretary Duncan has recognized the complexity and expense of building effective databases by including generous funding for states in the stimulus package. While a few states such as Texas and Florida are well advanced, it will be several years before most states will have the information to encourage richer evaluations of teachers and programs.

Reason #8: The rationale for tying tests to compensation is not clear.

One possible reason is to increase the effort, time, and resources devoted to teaching the content and skills to be tested. However, the consensus is very strong that the No Child Left Behind Act's testing mandate has narrowed instruction too much already at the expense of art, music, social studies, and foreign language instruction. A second reason might be to differentiate among teachers to identify the "slugs" from the "maestros." However, in most schools, one does not need a standardized test to identify the worst and best, and it will not work to sort those in the middle with sufficient precision to withstand the inevitable court challenge. A third reason might be to instill better practice. But if teacher compensation does not keep up with inflation because of poor student performance, then teachers will . . . what? Work harder? Dig deeper? Stay longer? There is no evidence that such measures improve instructional practices or student outcomes.

Secretary Duncan is correct when he catalogues the weaknesses in the present system of preparing, recruiting, mentoring, retaining, inspiring, retraining, promoting, and dismissing teachers. One of his main answers to all this—tying teacher compensation to standardized test results—is an idea that is way ahead of just about everything it would need to have even a chance of working fairly and reliably, if at all. He should revise the proposed regulations to give emphasis to other ideas—such as supporting high quality preschool that is tied to intensive early literacy in the primary grades—that we know can work today.

WRITTEN BY GORDON MACINNES, A FELLOW AT THE CENTURY FOUNDATION

OCTOBER 27, 2009

THE CENTURY FOUNDATION IS PUBLISHING THE ISSUE BRIEF SERIES TO HELP EXPLAIN AND CALL ATTENTION TO PUBLIC POLICY IDEAS THAT ARE WORTHY OF DISCUSSION AND DEBATE. THE VIEWS EXPRESSED IN THIS SERIES ARE SOLELY THOSE OF THE AUTHORS OF EACH ARTICLE.

THIS AND OTHER PUBLICATIONS CAN BE FOUND ON THE CENTURY FOUNDATION WEB SITE: www.tcf.org