

Connecticut Debate Association

February 12, 2011. Joel Barlow High School

Resolved: Teacher pay, promotion and tenure should be based primarily on measured student performance.

Teachers Lose Jobs Over Test Scores

Stephanie Banchero, Wall Street Journal (WSJ.com) July 24, 2010

Washington, D.C., Schools Chancellor Michelle Rhee fired 241 teachers Friday under a new evaluation system that holds teachers accountable for student test scores.

She also put an additional 17% on notice that if they don't improve next year, they could lose their jobs.

The move to oust 6% of the system's 4,300 teachers is the most drastic step yet in Ms. Rhee's aggressive bid to overhaul the work force of one of the nation's most troubled school systems. The district has been at the forefront of a national effort to rid classrooms of low-performing teachers. Of the 241 teachers fired, 165 were let go because of poor performance; the remainder didn't have proper teaching credentials.

Ms. Rhee said Friday she took over a system in 2007 where 95% of teachers were rated excellent and none terminated for poor performance. Yet, students posted dismal test scores. "The move today simply brings things into alignment," she said.

The head of the Washington Teacher's Union called the evaluation system flawed and said the union will challenge the firings.

"I'm not opposed to teachers being terminated, and I don't believe all 4,000 of ours are outstanding," said George Parker, president of the teacher's union. "But our teachers are entitled to an instrument that assesses their performance fairly, and this evaluation system does not."

Ms. Rhee has had an antagonistic relationship with the union. Friday's announcement further inflamed tensions and could factor into the hotly contested mayoral race. Incumbent Mayor Adrian Fenty hired Ms. Rhee and has promised to keep her on. His challenger, D.C. Council Chairman Vincent Gray, hasn't committed to keeping her.

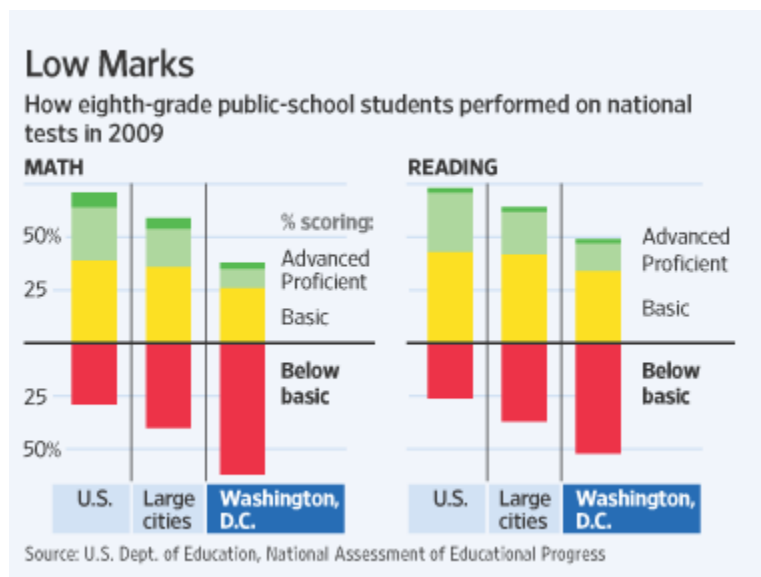
The teacher evaluation system developed under Ms. Rhee is one of the most rigorous in the nation. It requires numerous classroom observations of teacher performance and measures teachers against student achievement. It also allows Ms. Rhee to quickly get rid of poorly performing teachers. In most school

systems, dismissing teachers is time-consuming and costly.

Several recent studies have found that districts rarely fire tenured teachers no matter how bad their performance. A 2009 study of 12 urban, rural and suburban districts by the New Teacher Project found that half of the districts didn't dismiss even one teacher in the five years from 2003 to 2008, said the groups' president, Tim Daly.

"Superintendents across the country are going to take notice of this because she is doing major, major reform through relatively low cost," Mr. Daly said.

But critics contend that Ms. Rhee's



evaluation system was rushed into implementation before there was evidence to show it is an effective tool. High-stakes consequences, such as dismissal, shouldn't be tied to an unproven evaluation system, they argue.

Under the Washington, D.C., system, teachers are evaluated five times a year by school administrators and master teachers on such things as creating coherent lesson plans and engaging students. After an initial observation, teachers receive a plan detailing weaknesses and are offered coaching for improvement, district officials said.

For 20% of the work force—those who teach reading and math in 4th through 8th grades—student growth on state achievement exams counts for 50% of the evaluation.

District officials said they plan to broaden the student achievement component to high school teachers in future years.

Teachers are ranked into four categories. This year, 16% reached the highest ranking, compared with 45% in past years. Some 20% landed in the bottom rating, compared with 4% in years past.

Ms. Rhee said she couldn't break down how many teachers were dismissed due to low student achievement versus those dismissed for poor classroom performance.

About 730 teachers, some 17% of the work force, were identified as "minimally effective," the second-lowest rating. They won't receive the annual salary increase, and, if they don't show improvement next year, they will be dismissed, Ms. Rhee said.

"My hope is that many of them improve, but at the same time, we need to make sure the bar is high," she said. "I've got two children in the system, and I don't want a 'minimally effective teacher' and I don't think anyone else does, either."

Teachers to be graded on student test scores

By Jaime Sarrío The Atlanta Journal-Constitution

Hoping to attract and keep top teachers in public schools, Georgia is changing the way educators are hired, paid and rated through a new evaluation system with far greater emphasis on student performance.

The changes are spurred by the \$400 million Race to the Top grant, a program introduced by the Obama administration to jump-start school reform nationwide. Georgia won the grant in August; in return it pledged to rethink public school policy, including creating a new evaluation system for all teachers. For subjects where students take standardized tests, 50 percent of the teacher's performance would be based on their test scores. School leaders will also be judged by test scores when the new model rolls out in 26 districts this fall.

"We strongly believe that the most important thing in a student's education is the quality of the teacher in the classroom," said Erin Hames, who will oversee the plan's implementation as a deputy chief of staff for Gov.-elect Nathan Deal. "The heart of education improvement in Georgia has to be focused on the classroom and classroom teachers."

Less than 1 percent of Georgia's 143,000 educators and support staff received "unsatisfactory" ratings on annual performance evaluations in 2008-09, evidence the current system isn't working, according to state leaders. Evaluations currently take into account classroom observations, job training and interaction with students, but they do not focus enough on student achievement, they say. State law requires that academic gains are a factor in the evaluation process, but it doesn't specify how or to what degree.

Many educators agree that the current evaluation system is imperfect but have reservations about linking pay and performance to test scores. Some also question whether the state can develop and implement a fair system in a short period of time. Under the proposal, the changes will begin in 26 districts that agreed to take part in the Race to the Top grant, and they will gradually extend statewide within five years.

Tim Callahan, spokesman for the 80,000-member Professional Association of Georgia Educators, said the organization is OK with making student test scores a part of evaluations, but not 50 percent. The educator-advocacy group also has concerns about the lack of teacher input and the transition to new leadership both in the governor's office and the Department of Education.

"It has the potential to be a colossal failure," he said. "The resources aren't there to do the job well and do it right. It's like we're trying to go to the moon with a rubber band and a spit ball."

The basics of a new evaluation system — such as the link to student test scores — were laid out in Georgia's Race to the Top application. The state agreed to carry out those key reforms in exchange for federal money. But the details are still in the works and will be an ongoing process as the state introduces the changes first in the 26 districts, and later statewide.

Until now, local districts have used their own models to evaluate teachers, but most rate educators in two ways: satisfactory or unsatisfactory. Unsatisfactory ratings must be reported to the state. In 2008-09, 817 educators earned that classification, according to Department of Education officials, 0.6 percent of the total number of certified staff in public schools that year.

The new evaluations will have multiple categories so that educators have more feedback on their performance and a better understanding of training they need to improve. Teachers and principals will be rated on whether they reduced the achievement gap, on lesson plans or school improvement plans, and, for teachers, classroom observations.

But the most controversial and complicated change will be including student test data in teacher and principal evaluations. The new system will take into account "student-growth" scores, arrived at by plugging three years of a student's test results into a mathematical formula to predict future scores. Educators will be rated on how closely students follow that trajectory.

State and local officials are trying to decide which of the handful of growth models in use nationwide they will use. Complicated formulas are more accurate but harder to explain to teachers.

Tennessee, which has one of the oldest of these models in the country, now requires 50 percent of a teacher's evaluation to hinge on student scores. That's frustrated teachers who say they can't get a clear explanation of how the scores are calculated.

'A multitude of issues'

Georgia teachers are concerned about the impact of adding test scores to their evaluations.

Curtis Baxter, a seventh-grade social studies teacher in DeKalb County, said his annual evaluation normally consists of a 20-minute observation from a school administrator who later offers feedback on his teaching style. Student performance isn't a consideration.

Baxter, who has been an educator for 17 years, said he is not opposed to having test scores factor into his reviews, as long as teachers are given more autonomy to discipline students and manage their classrooms. Now, he says, students are sometimes promoted to higher grades regardless of their test scores, and sometimes returned to class even after serious behavior problems.

"There's a multitude of issues that need to be (talked) over before decisions are made," Baxter said.

One issue is how to evaluate the estimated 70 percent of Georgia teachers who instruct nontested classes such as physical education and kindergarten. According to the plan, 60 percent of evaluations for those teachers would be based on observations and walk-throughs, while the remainder would come from student and parent surveys.

The stakes will be high, with implications for teachers' continued employment and salaries.

After three years, new teachers whose students aren't showing enough academic growth on tests would not be recertified. Other teachers would be recertified every five years only if their students post the proper gains.

These changes have to be approved by the Professional Standards Commission, which oversees teacher certification. They are expected to take effect in five years, according to state officials.

Salary step increases, now determined by years of employment and education attainment, will be tied to performance for all teachers and leaders, and the top performers will earn extra pay increases. A new salary scale will take effect in the 26 districts in the fourth year of the grant, but it will require action by the General Assembly before it applies statewide.

Existing teachers will be able to choose whether to participate in the new pay system or continue in the existing salary schedule; but new teachers will be paid under the new system.

One Georgia school district, Jones County, backed out of the Race to the Top program because of objections to the performance pay plan. A high-profile study released this year by Vanderbilt University found that offering performance bonuses alone does not lead to higher test scores.

But Georgia Race to the Top architect Hames said she believes the changes will help weed out ineffective teachers and attract more high-quality instructors to the profession.

“You want to have a salary system that puts the focus on quality teaching, and helping students grow academically,” she said. “You want to reward those teachers who are wonderful at what they do.”

The state has pledged to include teachers in the development of the new system.

Nationally, teacher evaluations are gaining attention as districts around the country look for ways to improve the quality of the teaching work force, said Dan Goldhaber, director of the Center for Education Data & Research and a professor at the University of Washington Bothell.

Principals often hesitate to give out bad ratings because they have no control over a teacher’s salary or assignment, he said.

So instead, low-performing teachers are given satisfactory ratings and encouraged to find jobs in another school or district, a process known to educators as “passing the trash” or “the dance of the lemons.”

“There’s not a lot of evidence that alternative evaluation systems make a difference, but there’s a lot of evidence what we’re doing today isn’t effective,” Goldhaber said. “Teachers differ, but unless you document those differences, you can’t act on them.”

Teaching Math, Above and Beyond the Test

The New York Times, Room for Debate, Updated September 7, 2010, 03:26 PM

[Vern Williams](#) teaches honors math at Longfellow Middle School in Fairfax County, Va. He was named to the National Mathematics Advisory Panel in 2006.

I decided to become a math teacher when I was in ninth grade and never looked back. I managed somehow to evade most of the educational fads throughout the years, and I take pride in teaching serious mathematics in a traditional academically focused classroom.

I test my students and report their progress based on how they perform mathematically -- how well they actually do math problems. Even though I use objective testing measures when evaluating their progress, I understand that such tests, although carefully crafted for my particular students, do not necessarily paint a complete and clear picture. Since they also learn content and solve problems well above the level found on standardized tests, I expect them to max out on such tests.

Some of my best teaching occurs (and my students learn the most) when I present material that will never appear on any summative assessment. For gifted students especially, it is imperative to present to them mathematics that is both challenging and interesting. At times this may include assigning complex problems that may take days instead of minutes to solve. Never will such problems appear on state or national tests that are used to determine a teacher’s added value.

Since many of my seventh and eighth grade students have maxed out on standardized tests since fourth grade, I would probably be rated least effective based purely on added-value analysis. I would be outraged if such a label ended up in the newspaper.

I'm all for using objective tests to assess student progress and teacher effectiveness as opposed to using portfolios, self-esteem measures and other so-called authentic assessments. But even objective tests can be misused especially when classroom dynamics are ignored.

One student with severe emotional issues can change the entire social and academic character of a classroom. Such situations are rarely if ever explained when value-added results are reported. These results should therefore be used carefully as part of a teacher's evaluation when appropriate. If certain students have multiple teachers helping them with language skills, how will value-added results be reported? Who will be included in the report? In such cases, school results would seem to be a more appropriate measure of student success.

I suspect that most teachers would find value-added analysis to be problematic because versions of it are used without being fully validated. Yes, we all tend to know who the good teachers are but parents, students, teachers and administrators certainly take into account much more than just standardized test scores when describing those teachers.

If value-added analysis is to be used, it should be fully validated and other data should also be included when evaluating and labeling a teacher.

Protection From Favoritism

The New York Times, Room for Debate, Updated October 5, 2010, 10:22 AM

[Kevin Carey](#) is the policy director of Education Sector, a think tank, and a columnist for [The Chronicle of Higher Education](#).

Teachers tend to be suspicious of outside evaluation — and rightly so. The “drive-by” observation model used in most districts, which involves a few cursory classroom visits and the inevitable “satisfactory” rating, is a joke.

Giving principals greater discretion to determine raises and promotions opens the door to favoritism. Judging teachers by the number of students who pass No Child Left Behind-mandated tests is grossly unfair, penalizing those who take on the toughest assignments.

Value-added measures of teacher effectiveness can help solve these problems. Because they compare how much students learned during the year to how much they were expected to learn, based on how fast they learned in the past, they level the playing field for teachers who are assigned students of different ability. They protect teachers from favoritism by relying on hard numbers. They allow teachers with unorthodox methods to prove their worth.

No measure is perfect, of course. Value-added results should be interpreted carefully, in light of statistical margins of error. But perfection can't be the enemy of the good, and annual testing is here to stay.

We can no longer choose whether or not we should examine how much student test scores in a given teacher's classroom change over the course of a school year. We can only choose whether to use the best available statistical methods to interpret that information, and act on it fairly and responsibly.

There are a lot of great teachers in our public schools who labor in anonymity, receiving neither the recognition nor the compensation they deserve. Value-added measures, combined with peer evaluation and more rigorous classroom observations, can right this wrong.

Too Unreliable

The New York Times, Room for Debate, Updated September 7, 2010, 12:24 PM

[Linda Darling-Hammond](#) is the Charles E. Ducommun professor of education at Stanford University, where she is co-director of the Stanford Center on Opportunity Policy in Education. She was founding director of the National Commission on Teaching and America's Future, and she led President Obama's education policy transition team.

Teacher evaluation was a fly-by operation when I was a high school English teacher 30 years ago, and it has improved little in most districts since. So I understand why there is such enthusiasm for evaluating teachers based on their students' test score gains, now that such data are available.

Unfortunately, as useful as new value-added assessments are for large-scale research, studies repeatedly show that these measures are highly unstable for individual teachers. Among teachers who rank lowest in one year, fewer than a third remain at the bottom the next year, while just as many move to the top half. The top rankings are equally unstable. In fact, less than 20 percent of the variance in teachers' effectiveness ratings is predicted by their ratings the year before. This is why the National Research Council has said that this evaluation system "should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable."

The reasons are simple. Test score gains are caused by many variables in addition to the teacher: students' learning and language background, attendance, supports at home, previous and current teachers, tutors, curriculum materials, class sizes and other school resources. Out-of-school time matters too. Summer learning loss accounts for more than half the achievement differential between high- and low-income students. Thus, researchers have found that the very same teacher looks more "effective" when she is teaching more advantaged students -- and less effective when she teaches more students who are low-income, new English learners, or who have special education needs.

Tragically, evaluating and rewarding teachers primarily on the basis of state test score gains creates disincentives for teachers to take on struggling students, just as accountability systems that rate doctors on their patients' mortality rates have caused surgeons to turn away patients who are very ill. While scores may play a role in teacher evaluation, they need to be viewed in context, along with other evidence of the teacher's practice.

Better systems exist -- like the career ladder evaluations in Denver and Rochester, the Teacher Advancement Program and the rigorous performance assessments used for National Board Certification, all of which link evidence of student learning to what teachers do in teaching curriculum to specific students. These systems also help teachers improve their practice -- accomplishing what evaluation, ultimately, should be designed to do.

Valuable Feedback

The New York Times, Room for Debate, Updated September 7, 2010, 04:58 PM

[Amy Wilkins](#) is vice president of the [Education Trust](#), a research and advocacy group.

Parents know that some teachers are stronger than others. Teachers know it, too. Yet flawed and archaic evaluation systems in most school districts fail to make meaningful distinctions. In fact, [virtually every teacher in America is deemed to be doing at least a "satisfactory" job](#) -- even in schools where too few students are learning enough.

When teachers are denied honest information about how they are doing, how can they possibly be expected to improve?

Common sense says there must be a better way to evaluate teacher performance. And there is. [Years of research](#) suggest that, while not perfect, "value-added" methods provide a far more honest and fair picture of how well teachers are doing, especially when coupled with rigorous classroom observation.

[This method measures](#) the contribution of an individual teacher to an individual student's learning. Instead of relying on a single end-of-the-year test score, it examines growth over the course of a school year. So even when a student enters a classroom far below grade level, if that student makes big learning

gains, the teacher gets credit for those gains. In fact, she gets far more credit for that student than for one who started the year a little above average but ended in the same place.

When summed over several years, these data can provide teachers with valuable feedback about what kinds of students they are most successful with and with whom they need to improve. They can help schools match the most able teachers with the students who most need them. And they can help leaders better target teacher supports and rewards. But relatively few districts actually use their data in these ways. And that's unfair to our teachers and their students.

No one is suggesting that "value-added" measures be the sole criteria of teacher reviews. But they provide important information that teachers and principals need. Indeed, when presented with their data last month, Los Angeles teachers [commonly expressed frustration that the district hadn't shared it sooner](#). To continue to ignore their frustration would be a terrible and costly mistake.

A Big Margin for Error

The New York Times, Room for Debate, Updated September 7, 2010, 12:30 PM

[Diane Ravitch](#) is the author of "The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education."

The United States Department of Education wants school districts to use "value-added" assessments to judge teacher quality. But this method is problematic.

Federal officials expect this methodology to identify "good" and "bad" teachers, making it possible to reward the good ones and fire the bad ones. If enough "bad" teachers are fired, goes the argument, test scores would soar, and the achievement gap between different races would close.

[Testing experts agree that the method is too error-ridden](#) to use to decide which teachers are rewarded or fired.

In July, research published by the U.S. Department of Education found that this methodology produces error rates of 25 percent based on three years of student scores; with fewer years, the error rate is even higher. Using it, many good teachers will wrongly be identified as "ineffective," and many less successful teachers will wrongly be labeled "effective."

A teacher who is effective one year may not be effective the next year, not because her "quality" changed, but because of the composition of her classes. Almost 70 percent of teachers who rank at the top one year may not be at the top the following year.

Student scores are influenced by many factors other than the teacher. Students are not randomly assigned to teachers. Scores may be depressed by student absenteeism, and they may be elevated by after-school tutoring. Many classes have more than one teacher, and in some districts, student mobility makes it hard to know which teacher should get credit or blame.

Proponents of the value-added method assume that the tests themselves are unerring. As New Yorkers learned recently, state test scores were vastly inflated over several years and produced inaccurate data.

The spread of this way of evaluating teachers is likely to cause negative consequences. Raising the stakes for test scores will promote teaching to the tests and cheating. It is also likely to narrow the curriculum only to what is tested: reading and math. There will be less time for the arts, history, geography, civics, science, foreign languages, even physical education.

The best teacher evaluation depends on the wise judgment of experienced supervisors, who consider a variety of factors, including observations of the teacher's work, the quality of student work and peer review. There is no technocratic fix for the problems of American education.

More Harm Than Good

The New York Times, Room for Debate, Updated September 7, 2010, 12:31 PM

[Jesse Rothstein](#) is an associate professor of public policy and economics at the University of California, Berkeley. He has studied the relationship between classroom assignments and estimates of "value-added" by teachers.

It is widely agreed that underperforming teachers should be identified early, helped to improve and, if necessary, removed from the classroom. But teacher evaluation is not as simple as many reformers assume, and "value-added" assessments are, at most, a small part. Rushing ahead with an experimental approach is likely to do more harm than good.

This type of assessment has a number of important shortcomings, ably reviewed in this [recent report](#). My own research has focused on two. First, it can only accurately measure teacher effectiveness if students are randomly assigned to teachers, but good principals carefully match students to the teachers best suited for their particular personalities and needs. The assessment models then penalize teachers working with hard-to-teach students while rewarding those with easier assignments.

Second, value-added analysis does a better job of measuring teachers' temporary effects than their longer-run contributions. Students with high-value-added third grade teachers earn much higher third grade standardized test scores than those with lower-value-added teachers that year. But this advantage dissipates very quickly, with more than two-thirds lost within one year and nearly all of the rest gone after two years. Can it be that effective teaching just doesn't matter for students' long-run achievement? Almost certainly not. More likely, truly effective teaching does have long-run payoffs, but value-added analysis does a poor job of measuring it.

Proponents of this type of assessment argue that it is self-evidently better than the current system. But such a system could well be worse than the status quo. Pressure to raise their scores will lead teachers to neglect non-tested topics and subjects, to drill test-taking skills rather than pursuing deeper learning with greater long-term payoff, and to avoid troubled students and others who may depress measured performance.

We need evaluation systems that limit these skewed incentives. To design them, we might look to private-sector evaluations of workers engaged in complex, multifaceted tasks. These rarely rely on simple quantitative measures. Rather, employers devote great effort and expense to collecting the subjective assessments of well-informed managers and coworkers. This would be difficult for teachers, who generally work alone and are formally observed in the classroom only a handful of times each year. To replace these haphazard observations with more frequent visits from trained evaluators and master teachers will require substantial additional resources. But it is necessary for serious assessments of teacher effectiveness. Cutting corners could easily hurt students more than it helps.

Imperfect, but Useful

The New York Times, Room for Debate, Updated November 10, 2010, 01:39 PM

[Marcus Winters](#) is a senior fellow at the Manhattan Institute, where he has done several studies on education testing and school report cards.

An objective measure of what an individual teacher adds to a student's proficiency.

As a statistical tool, "value-added" assessment is not a perfect measure of teacher quality. Indeed, there will never be a perfect measure of teacher quality. The real question, then, is whether this analysis can improve the methods we use to evaluate teacher performance. There is ample reason to believe that it can.

Currently, nearly all teachers are rated satisfactory or higher. No one honestly believes that this system accurately assesses teacher quality. Uniform satisfactory ratings are inconsistent with empirical research showing wide variation in teacher quality, and they defy common sense.

When evaluating teacher quality, what we really want to know is: To what extent does an individual teacher contribute to her student's proficiency? Value-added test-score analysis provides us with an objective estimate of this contribution.

It is true that the test-score analysis doesn't tell us everything we want to know about a teacher. For example, if a teacher is performing poorly, it can't tell us why. Also, random error ensures some teachers will receive ratings that over- or underestimate their actual influence on student proficiency. For those and other compelling reasons, it would be irresponsible to use only value-added analysis to evaluate teachers. Nonetheless, imperfect value-added assessment is surely an improvement upon the current system, which makes no meaningful attempt to differentiate teachers by their effectiveness. Further, recent empirical research finds evidence that value-added measures of a teacher's past performance are far better predictors of her future students' achievement than are other observed characteristics, like her credentials and years of experience.

Everyone agrees that teacher quality is crucial for student learning. It's well past time that public schools took seriously the job of identifying it. Value-added assessment is an important and powerful tool that public schools should make use of to ensure that the most effective teachers are educating our kids .

Test scores can't prove whether teacher experience matters

The Washington Post, The Answer Sheet by Valerie Strauss, 12/23/2010

This was written by Monty Neill, interim executive director at The National Center for Fair & Open Testing, known as FairTest.

By Monty Neill

According to Bill Gates and Education Secretary Arne Duncan, teacher experience does not matter very much. Duncan says we should not "pay teachers based on their educational credentials and seniority" -- that is, knowledge and experience. Gates also recently denigrated teacher experience and masters' degrees. He cited Jennifer R King's literature review, which shows that experience does matter, but mostly that which they gain in their first five years of teaching.

But here's the rub: They rely entirely on standardized test scores for evidence, even though the tests fail to provide adequate evidence for drawing these conclusions.

Even Duncan agrees that currently existing state tests over-emphasize basic skills and do not assess many essential aspects of learning. (Duncan's proposed "solution," the multi-state consortia tests, is not likely to help very much, nor will "little value to the addition" uses of those tests.)

The tests are beatable with test prep schemes, for which a teacher does not need deep subject knowledge or an understanding of how each of her students best learns. A revolving door of minimally-trained, low-paid new teachers delivering from a script may be able to boost scores on low-level tests nearly as well as more experienced teachers. Is that what we want from our schools? Is it OK if it mostly happens to low-income children, where teaching to the test is most common?

Surveys have found that parents, communities, even legislators want far more from their schools than only academics -- never mind academics reduced to test prep. Their goals for schools include basic skills, critical thinking, arts and literature, preparation for skilled work, social skills, citizenship, and physical and emotional health. (See Richard Rothstein's Grading Education, Ch 2, for one example). Teachers must help our children develop these areas of knowledge, skill and competence.

Unless the U.S. educational goal is merely to boost scores on low-level tests, we need answers to more important questions before we evaluate the worth of teacher experience or degrees.

These questions include: Are teachers with more experience better than novices at ensuring deep understanding, critical thinking, and the ability to apply knowledge? Do their capacities grow beyond the first five or ten years of experience?

Do more experienced teachers help students acquire the full range of outcomes Rothstein listed? Do they work more effectively toward comprehensive education goals with students in poverty, those with disabilities or limited English proficiency? Do they better help students become more engaged with their

learning? Are they more successful in classrooms with a great deal of social, cultural and economic diversity?

Can they better provide the ‘soft’ skills that research increasingly finds are essential to later success in life? I expect in many cases the answer would be yes, significantly so.

Since test scores don’t tell us what we need to know, we need studies that look carefully at actual student work, as well as the long-term results of schooling. This can be done, but policymakers have shied from doing it, content to rely on inexpensive low-level tests.

Finally, the attacks on experienced teachers seem more motivated by politics and budgets than by research evidence. In a period of sharp budget cuts, the claim that teacher experience does not matter is increasingly used to justify the hiring of under-prepared novices to replace experienced teachers. The novices are expected to use scripted curricula to train, not educate, their mostly low-income and minority-group students, in order to boost test scores. If test scores rise like hot air balloons, that will be presented as “evidence” of success.

These issues raise broader questions about the goal of schooling. Is it to be merely employment driven functionality, for many students just to fit requirements of jobs with little cognitive demand? Or does this nation want educated students who can participate thoughtfully in civic and social issues and be lifelong learners? Those questions are not technical and cannot be answered by changes in standardized test scores. They require the consideration and actions of an involved citizenry.

Method to Grade Teachers Provokes Battles

By SAM DILLON, The New York Times, August 31, 2010

How good is one teacher compared with another?

A growing number of school districts have adopted a system called value-added modeling to answer that question, provoking battles from Washington to Los Angeles — with some saying it is an effective method for increasing teacher accountability, and others arguing that it can give an inaccurate picture of teachers’ work.

The system calculates the value teachers add to their students’ achievement, based on changes in test scores from year to year and how the students perform compared with others in their grade.

People who analyze the data, making a few statistical assumptions, can produce a list ranking teachers from best to worst.

Use of value-added modeling is exploding nationwide. Hundreds of school systems, including those in Chicago, New York and Washington, are already using it to measure the performance of schools or teachers. Many more are expected to join them, partly because the Obama administration has prodded states and districts to develop more effective teacher-evaluation systems than traditional classroom observation by administrators.

Though the value-added method is often used to help educators improve their classroom teaching, it has also been a factor in deciding who receives bonuses, how much they are and even who gets fired.

Michelle A. Rhee, the schools chancellor in Washington, fired about 25 teachers this summer after they rated poorly in evaluations based in part on a value-added analysis of scores.

And 6,000 elementary school teachers in Los Angeles have found themselves under scrutiny this summer after The Los Angeles Times published a [series of articles](#) about their performance, including a searchable database on its Web site that rates them from least effective to most effective. The teachers’ union has protested, urging a boycott of the paper.

Education Secretary [Arne Duncan](#) weighed in to support the newspaper’s work, calling it an exercise in healthy transparency. In a speech last week, though, he qualified that support, noting that he had never released to news media similar information on teachers when he was the Chicago schools superintendent.

“There are real issues and competing priorities and values that we must work through together — balancing transparency, privacy, fairness and respect for teachers,” Mr. Duncan said. On The Los Angeles Times’s publication of the teacher data, he added, “I don’t advocate that approach for other districts.”

A [report](#) released this month by several education researchers warned that the value-added methodology can be unreliable.

“If these teachers were measured in a different year, or a different model were used, the rankings might bounce around quite a bit,” said Edward Haertel, a Stanford professor who was a co-author of the report. “People are going to treat these scores as if they were reflections on the effectiveness of the teachers without any appreciation of how unstable they are.”

Other experts disagree.

William L. Sanders, a senior research manager for a North Carolina company, SAS, that does value-added estimates for districts in North Carolina, Tennessee and other states, said that “if you use rigorous, robust methods and surround them with safeguards, you can reliably distinguish highly effective teachers from average teachers and from ineffective teachers.”

Dr. Sanders helped develop value-added methods to evaluate teachers in Tennessee in the 1990s. Their use spread after the 2002 [No Child Left Behind](#) law required states to test in third to eighth grades every year, giving school districts mountains of test data that are the raw material for value-added analysis.

In value-added modeling, researchers use students’ scores on state tests administered at the end of third grade, for instance, to predict how they are likely to score on state tests at the end of fourth grade.

A student whose third-grade scores were higher than 60 percent of peers statewide is predicted to score higher than 60 percent of fourth graders a year later.

If, when actually taking the state tests at the end of fourth grade, the student scores higher than 70 percent of fourth graders, the leap in achievement represents the value the fourth-grade teacher added.

Even critics acknowledge that the method can be more accurate for rating schools than the system now required by federal law, which compares test scores of succeeding classes, for instance this year’s fifth graders with last year’s fifth graders.

But when the method is used to evaluate individual teachers, many factors can lead to inaccuracies. Different people crunching the numbers can get different results, said Douglas N. Harris, an education professor at the University of Wisconsin, Madison. For example, two analysts might rank teachers in a district differently if one analyst took into account certain student characteristics, like which students were eligible for free lunch, and the other did not.

Millions of students change classes or schools each year, so teachers can be evaluated on the performance of students they have taught only briefly, after students’ records were linked to them in the fall.

In many schools, students receive instruction from multiple teachers, or from after-school tutors, making it difficult to attribute learning gains to a specific instructor. Another problem is known as the ceiling effect. Advanced students can score so highly one year that standardized state tests are not sensitive enough to measure their learning gains a year later.

In Houston, a district that uses value-added methods to allocate teacher bonuses, Darilyn Krieger said she had seen the ceiling effect as a physics teacher at Carnegie Vanguard High School.

“My kids come in at a very high level of competence,” Ms. Krieger said.

After she teaches them for a year, most score highly on a state science test but show little gains, so her bonus is often small compared with those of other teachers, she said.

The Houston Chronicle reports teacher bonuses each year in a [database](#), and readers view the size of the bonus as an indicator of teacher effectiveness, Ms. Krieger said.

“I have students in class ask me why I didn’t earn a higher bonus,” Ms. Krieger said. “I say: ‘Because the system decided I wasn’t doing a good enough job. But the system is flawed.’ ”

This year, the federal Department of Education’s own research arm warned in a study that value-added estimates “are subject to a considerable degree of random error.”

And last October, the Board on Testing and Assessments of the National Academies, a panel of 13 researchers led by Dr. Haertel, wrote to Mr. Duncan warning of “significant concerns” that the Race to the Top grant competition was placing “too much emphasis on measures of growth in student achievement that have not yet been adequately studied for the purposes of evaluating teachers and principals.”

“Value-added methodologies should be used only after careful consideration of their appropriateness for the data that are available, and if used, should be subjected to rigorous evaluation,” the panel wrote. “At present, the best use of VAM techniques is in closely studied pilot projects.”

Despite those warnings, the Department of Education made states with laws prohibiting linkages between student data and teachers ineligible to compete in Race to the Top, and it designed its scoring system to reward states that use value-added calculations in teacher evaluations.

“I’m uncomfortable with how fast a number of states are moving to develop teacher-evaluation systems that will make important decisions about teachers based on value-added results,” said Robert L. Linn, a testing expert who is an emeritus professor at the University of Colorado, Boulder.

“They haven’t taken caution into account as much as they need to,” Professor Linn said.

Value-added modeling

From Wikipedia, the free encyclopedia

Value-added modeling (also known as **value-added analysis** and **value-added assessment**) is a method of teacher evaluation that measures the teacher's contribution in a given year by comparing current school year test scores of their students to the scores of those same students in the previous school year, as well as to the scores of other students in the same grade. In this manner, value-added modeling seeks to isolate the contribution that each teacher makes in a given year, which can be compared to the performance measures of other teachers.^[1]

Method

Statisticians use a student's past test scores to predict the student's future test scores, on the assumption that students usually score approximately as well each year as they have in past years. The student's actual score is then compared to the predicted score. The difference between the predicted and actual scores, if any, is assumed to be due to the teacher and the school, rather than to the student's natural ability or socioeconomic circumstances.

In this way, value-added modeling isolates the teacher's contributions from factors outside the teacher's control that are known to strongly affect student test performance, including the student's general **intelligence**, **poverty**, and parental involvement.

By aggregating all of these individual results, statisticians can determine how much a given teacher typically improves student achievement, compared to how much the typical teacher would have improved student achievement.
