




*Listening. Learning. Leading.*

A stylized, blue-tinted American flag with white stars and stripes, set against a dark blue circular background. The flag is draped and folded, creating a sense of movement and depth.

EDUCATION  
ISSUES  
2007



Kurt Landgraf  
President & CEO

It has been about five years since President Bush signed into law the No Child Left Behind Act, setting the stage for dramatic and far-reaching education reform based on standards and accountability.

In enacting this landmark legislation, the Congress and the President acted with both passion and pragmatism. There has been bipartisan support to make the law a reality in every classroom in America. Real progress is being made across the nation as we continue to improve teaching and learning for every child. We are on the road to fulfilling the promise of the No Child Left Behind Act, but much work remains.

Continuing in our mission to advance quality and equity in education, ETS has prepared a briefing book on education issues in order to provide useful information to elected and appointed officials at the local, state and federal levels. It offers a balanced picture of three critical education issues:

- accountability and the role of assessments in providing a window into the performance of our schools
- schools' ability to attract and retain highly qualified teachers
- progress in closing the education achievement gap

ETS has been working to help America's schools take advantage of the opportunity for education reform. We expect to do even more. We believe in the children of this great nation and their right to an excellent education. And we hope that you will find the material in these pages informative and helpful.

Sincerely,

A handwritten signature in cursive script that reads "Kurt M. Landgraf".

Kurt Landgraf  
President & CEO  
ETS



# Education Issues 2007

<b>Introduction</b>	.....	5
<b>Section I. Accountability and Assessment</b>	.....	6
Basics of Assessments	.....	8
Education Week's Quality Counts at 10	.....	11
The No Child Left Behind Act (NCLB) Overview	.....	14
ETS on the Issues	.....	15
<b>Section II. Teacher Quality</b>	.....	17
Where We Stand on Teacher Quality	.....	18
A Primer on Value-Added Models	.....	29
ETS on the Issues	.....	41
<b>Section III. Access and Equity</b>	.....	42
One-Third of a Nation	.....	44
ETS on the Issues	.....	47
<b>Section IV. The Public Speaks on Education and Competitiveness</b>	.....	49
<b>Section V. Student Performance Data</b>	.....	53



# Introduction



The quality of public education is important to our nation because it shapes the lives of so many individuals and families, as well as our society as a whole. With its focus on standards and accountability, the No Child Left Behind (NCLB) Law set the stage for historic education reform.

While there is certainly debate over how to accomplish education reform, the need to improve public schools has near-universal support. There has been a bipartisan effort to make the law a reality in every classroom in every school district in America. Real progress is being made across the nation as we continue to improve teaching and learning for every child. We are on the road to fulfilling the goal, but much work remains.

We at ETS are deeply involved in American education. As the country's largest private, nonprofit educational testing and research company, ETS assesses student learning at all levels. We develop, administer or score 24 million tests annually in more than 180 countries in pursuit of our mission. Our tests license professionals in a variety of occupations. In our public policy reports, we strive to present an accurate picture of the opportunities and challenges facing American education. And as a leader in education research, we analyze issues such as fairness in assessments, the impact of technology on learning, how to motivate students, and the characteristics of good teaching. We inform and help to guide policymakers on the development, use and interpretation of test results.

This briefing book on education issues seeks to present useful information to elected and appointed officials at the local, state and federal levels. It offers a balanced picture of three critical education issues:

- accountability and the role of assessments in providing a window into the performance of our schools
- schools' ability to attract and retain highly qualified teachers
- progress in closing the education achievement gap

ETS is also making its staff of experts available to policymakers, reporters, parents and other concerned citizens who have questions about the issues raised in this report.

Our mission at ETS is to advance quality and equity in education worldwide. We are committed to improving instruction and learning in America's schools, and we hope that you will find the information in these pages informative and helpful.

We invite you to continue this discussion of these important issues at [www.ets.org/letstalk](http://www.ets.org/letstalk).



# Accountability and Assessment: A Window Into Education

## **No Child Left Behind Is Transforming Public Education**

---

The No Child Left Behind Act (NCLB) is a forceful challenge to American education. Signed into law in January 2002, NCLB strengthens the federal government's role in elementary and secondary education. The law reforms education through mandates on standards, and then holds the schools accountable through assessments. States are required to adopt specific plans to comply with the law's mandates. School districts are to be held to uniformly high standards of excellence. And students will be monitored by standardized tests given annually in grades three through eight in reading, math and science.

## **For Reform to Work, Assessments Must Work**

---

For schools and students to be held accountable, assessments must be accurate tools. Good tests must be reliable, valid and fair. Tests must measure the academic standards they are supposed to measure; they must produce scores that are consistent among groups of students; and they must fairly reflect differences in knowledge, not in gender or ethnic backgrounds. If properly developed and administered, assessments will not only monitor student achievement, they will also help education leaders focus resources and shape reform.

## **Assessments Support Learning; They Don't Punish Students**

---

Testing is just one step in education reform. Tests provide answers about students' strengths and weaknesses, so that teachers can make better choices. Tests help parents understand how well their children are mastering their state's curriculum compared with other students. Tests allow parents to make important decisions about their children's education, including whether their schools are meeting the NCLB requirements for Adequate Yearly Progress (AYP). And good tests help schools determine their own performance compared with that of other schools. Tests should not penalize. Tests act as a supportive tool, allowing policymakers, administrators and teachers to inform instruction and improve learning. For tests to work, we must all respond to their results — particularly poor results — thoughtfully and responsibly to improve classroom achievement.

## ACCOUNTABILITY & ASSESSMENT

---

*Now, I've heard them say, well, tests, we're testing too much. If you don't like to take tests, too bad. Because we need to know, we need to know whether you're learning.*

— President George W. Bush

*When you are sick, the doctor performs a series of tests to determine what illness you have. Then the doctor prescribes a remedy specific to the illness. The academic tests under this bill serve the same purpose — they help teachers and parents diagnose the problems, and apply remedies that will help each child to learn more effectively.*

— U.S. Sen. Edward M. Kennedy, D-Mass.

*Among political leaders, testing is turning into a means of reform rather than just a way of finding out whether reforms have been effective.*

— Paul E. Barton, Senior Associate,  
Policy Information Center, ETS

*We're seeing progress. Let me tell you the facts. The test scores for fourth-grade math around the nation went up 9 points between the years 2000 and 2003. See, I'm able to tell you that because we measure. If we hadn't measured, I couldn't tell you that. Eighth-graders improved by 5 points in the same period. In other words, math scores are beginning to go up. Still too low, but they're heading toward higher standards. Reading scores for fourth-graders increased in the vast majority of states that tested between 1998 and 2003, including the state of Arkansas. See, your reading scores are going up here.*

— President George W. Bush, speaking in  
Van Buren, Ark.

*Our school got the outstanding achievement this year. Last year we were a low score. This year it went up. We are really proud of that. We have a lot to deal with — all the teachers that are here working together as a family school staff. I think all the students are into it already, and I think that's why we got the recognition.*

— Second-grade teacher, New York, N.Y.,  
ETS focus group for "Equity and Adequacy:  
Americans Speak on Public School Funding,"  
a public opinion poll conducted in 2004 for ETS

*Today my third-graders just took a state test and that state test — that test was so, so difficult, I mean it was a killer, but 99 percent of my children passed that — my third-graders passed that test with 61 percent of them being on free and reduced lunch, because we are about advancing the learning, moving them up a notch, rather than the status quo.*

— Third-grade teacher, Dallas, Texas,  
ETS focus group for "Equity and Adequacy:  
Americans Speak on Public School Funding,"  
a public opinion poll conducted  
in 2004 for ETS

*For years in this country we accepted mediocrity and failure from students and finally we said, "enough — we can and must do better." The No Child Left Behind Act has brought substantial changes which haven't been easy but which, we now see, have spurred real progress. But we must stay with our commitment to the law and to our schools to continue that progress until all students are proficient at reading and math.*

—U.S. Rep. George Miller, D-Calif., Ranking  
Member of the House Education and Workforce  
Committee, July 2005

# BASICS OF ASSESSMENTS

---

Tests are needed to assess students and schools. They help improve teaching and learning. And they generate information that is vital to making policy decisions.

Yet creating tests of the highest quality and fairness is a complex process. The process of creating good tests includes setting clear objectives and purposes, selecting and preparing questions, pretesting, obtaining test reviews, and continuously reviewing the performance of the assessment test developers.

The quality of tests is vitally important to the quality of the education system we serve.

## The Purpose of Assessments

---

There are many kinds of assessments that serve a variety of purposes. Two purposes of standardized tests are for use in admission to college or graduate school and for placement in programs. Licensure tests are used to protect the public from harm by helping to ensure that people who want to drive cars, sell insurance, perform surgery, and so forth have the necessary knowledge and skills. Teachers use tests to track students' progress in mastering curriculum, for instructional management, and to report progress to parents, school leaders and public officials. A single assessment is merely a snapshot of a particular student on a particular day.

Most states have adopted standards specifying the content and performance level required of students in grades K-12. Curricula and assessments have been adopted in most states to specify and measure student achievement of the state's curriculum standards. These standards-based assessments are the primary focus of this document because most states and school districts are in the process of implementing standards-based curricula and assessments. Furthermore, the No Child Left Behind Act (NCLB) ensures that standards-based assessments will be at the forefront of the education policy debate for a number of years.

Assessments are also used for accountability. There has been a marked increase in that type of assessment over the past two decades. The NCLB Act requires annual testing of students in reading and mathematics in grades three through eight. States must adopt standards in science and assess science achievement beginning in 2008. The implementation

of that law provides an opportunity to re-examine assessments and their purposes, and to consider ways of streamlining the assessment process.

A single test can be used for more than one purpose, but trying to accomplish too many purposes with a single test can result in confusion about what is being measured and why, and the appropriate use of the test score.

## What Are the Characteristics of Good Assessments?

---

There can be no such thing as a test that is perfect in all respects. Every test is necessarily a compromise among competing demands. For example, educators may wish to test everything that was taught in math during the school year, but a test that covers everything may take many days to administer and consume too much instructional time. Those in charge of the testing program must reach a compromise between adequate content coverage and acceptable administration time.

Whatever compromises are made, good assessments have certain characteristics in common. They must be reliable, valid, fair and affordable.

**Reliable.** If scores are reliable, they are not greatly affected by chance factors, such as the particular person who scored an answer to an essay question, or the particular questions that happened to be used on different versions of the same test. An unreliable score is like a measurement made with a ruler that gives different lengths for the same table each time you measure it. Scores on unreliable tests cannot be trusted because they contain lots of random wobble.

The extent of random wobble in a test score is indicated by a statistic called the “standard error of measurement.” Reputable test publishers will describe the standard errors of measurement for their tests. When a test score is used to make high-stakes decisions about an individual, it is a good idea to treat the score as a range that extends one standard error above and one standard error below the particular score a person obtained.

If scores are reported for groups of test takers, reliability is much less of a concern because the negative random wobbles in some scores will be balanced by positive random wobbles in other scores, and the group average will be much more stable than any of the individual scores used to calculate the average.

**Valid.** Validity is the extent to which the inferences based on test scores are appropriate and supported by evidence. Validity is the single most important characteristic of any test. A good way to think about validity is to consider the kinds of evidence it would take to convince you that the test is doing its job properly.

For standards-based tests, it is important to show that the questions really are measuring the content they were intended to assess. Do experts agree that the questions are strongly linked to the standards to be measured? Do the different parts of the test relate to one another in reasonable ways? For example, do the math questions relate more closely to one another than they do to the reading questions? Do the scores agree with school-based measures of success? For example, do students with good grades in school tend to score higher than students with poor grades?

Look for evidence that the claims made about the results of using the test are becoming reality. For example, if one of the claims is that use of the test will narrow the gap between minority and majority performance, do any data show the gap is really being narrowed? Be alert to unintended negative consequences such as an increase in teacher turnover or an increase in dropout rates that can be attributed to the use of the test.

No single number will show the validity of a test, but the people sponsoring the test should be able to show evidence that use of the test is appropriate.

**Fairness.** One of the most controversial aspects of testing is fairness. A major reason for the controversy is that some people believe tests are not fair as long as different groups obtain different average scores.

Group differences are a sign that there may be a problem with fairness, but differences alone are not proof of unfairness, because there may be a real and relevant difference between groups in knowledge of what the test is supposed to be measuring. The average height of men differs from the average height of women, but that does not mean yardsticks are unfair. Tests that cause differences to appear because of factors unrelated to the purpose of the test are not fair. For example, group differences caused by unnecessarily difficult vocabulary in a math test are not fair because reading is not related to the purpose of the test.

Tests that correctly reflect real differences in knowledge or skill related to the purpose of the test are fair.

**Affordable.** Resources for education are scarce and must not be wasted. Tests should use no more class time than is necessary and should cost no more than is necessary to produce scores that are reliable, valid and fair.

Affordability requires compromises. For example, educators may desire a science assessment that requires students to actually design and carry out experiments. There are some advantages to such a performance-based assessment. Policymakers must decide, however, if the advantages of performance-based assessments over multiple-choice tests that cover the same content are worth the additional investment in class time for test administration and the additional costs of observing and scoring the students' actions.

Good testing organizations can produce high-quality, affordable tests, but they cannot produce ideal tests that will satisfy all constituencies at an affordable cost. Even good, reliable, valid and fair

testing programs, therefore, are likely to have many critics who disagree with the compromises that were made. As a result, testing is likely to remain the subject of intense policy debate.

### **The Role of Standards-Based Assessment in Education**

---

Tests are a crucial component of standards-based educational reform, but tests alone are not sufficient. Tests by themselves can no more improve learning than thermometers can cure patients. Tests and thermometers provide useful information, but structures must be in place to allow appropriate use of the information before it can do any good. To ensure that tests have a beneficial impact, they must be part of an integrated system that aligns standards, curricula, tests and uses of the scores.

### **Before Any Standardized Testing Is Done, There Should Be:**

---

- clear, specific and generally accepted content standards indicating what students should know and be able to do
- curricula and materials that are closely aligned with the standards
- training for teachers in appropriate methods for standards-based education
- adequate notice given to students and their parents or guardians that new requirements for promotion or graduation will be put in place
- adequate opportunity for students to learn the material that will be tested and an opportunity to become familiar with the test format that will be used

### **Tests Used in Standards-Based Reform Should Be:**

---

- clearly and closely linked to the standards in the content and skills that are measured
- accepted as fair by reviewers who reflect the diversity of the test takers and the various constituencies of the community
- available in versions suitable for students with disabilities
- sufficiently reliable and valid to support decisions that affect students' lives
- given in surroundings that mitigate stress (e.g., comfortable seats, reasonable temperature, quiet, breaks when appropriate)

### **After the Testing Is Done, There Should Be:**

---

- meaningful remedial learning experiences for students who performed poorly
- multiple opportunities for students who performed poorly to take different forms of the test
- uses of aggregated data to determine necessary changes in curricula, materials or teaching methods
- understandable reporting to the public of the meaning of the results
- periodic review of the tests, standards, curricula, materials and teaching methods, to ensure all remain current and aligned with one another

# EDUCATION WEEK'S QUALITY COUNTS AT 10: A DECADE OF STANDARDS-BASED EDUCATION

---

## Executive Summary: Below the Surface

---

An analysis of test data finds the results both heartening and sobering.

State efforts to carry out standards-based education over the past decade have a positive, but modest, relationship with gains in student achievement on the National Assessment of Educational Progress (NAEP), according to this year's edition of *Quality Counts*.

The 10th-annual report by *Education Week* examines the progress states have made on a core set of policy indicators related to standards-based education over the past decade. It also explores the relationship between changes on those indicators and gains in student learning, as measured by NAEP reading and mathematics tests in grades 4 and 8 from 1996 to 2005.

The report does not examine the more recent impact of the federal No Child Left Behind Act, which President Bush signed into law in 2002.

The results are at once heartening and sobering. They're heartening because when looked at over more than a decade, student achievement has gotten better, particularly in mathematics and particularly for low-income and minority students. An increasing number of states also have embraced a standards-based-education framework, with some of the earliest and most ardent adopters of standards-based accountability systems making some of the most progress in student achievement.

Nationally, NAEP scores in fourth-grade math have increased by 18.5 points on a 500-point scale, or nearly two grade levels, since 1992, near the start of the standards movement.

Even more encouraging are the gains for Black and Hispanic fourth-graders: 27.7 points and 24.2 points, respectively. One way to think about those gains is that if the scores for White students had not also improved, the advances would have been enough to shrink the Black-White achievement gap that existed in 1992 by 80 percent, and the gap between Hispanics and non-Hispanic Whites by 94 percent, virtually closing the gap between those groups in fourth-grade math.

The scores for low-income students, which NAEP began reporting in 1996, basically mirror the average national increases.

In math, improvement has come at the "basic" and "proficient" levels on NAEP, both for students overall and for Black, Hispanic, and low-income students. Those gains are mirrored across most states, according to special analyses of NAEP scale scores conducted for *Quality Counts 2006* by ETS, which looked at changes in state performance between 1992 and 2005. Of the 41 states and the District of Columbia with data, all made significant progress in fourth-grade math achievement, and all but two states made significant progress in eighth-grade math achievement during that period.

It's hard to be as sanguine about reading. The national average in reading barely budged from 1992 to 2005, inching up just 2 points in both grades 4 and 8. But, even here, somewhat better news lies beneath the surface. The scores for Black, Hispanic, and low-income students in fourth-grade reading increased at nearly triple the national average, or about two-thirds of a grade level.

Nationally, the achievement gaps narrowed somewhat for all groups in reading and math and at both grade levels — except between Black and

White students in eighth-grade reading. The gap-closing was significant between Black and White students in math at both grades, and between Hispanic and non-Hispanic White students in fourth-grade math.

Progress is even more mixed at the state level, according to the ETS analysis, although the picture is complicated by the fact that many states either did not take part in the state-level NAEP during the periods examined or did not have enough poor or minority students in the NAEP samples to permit valid comparisons of change over time.

The 2006 report highlights individual states — including Delaware, Massachusetts, New York, North Carolina and Texas — whose progress stands out over the past decade, with in-depth profiles that explore what might explain such changes.

The more controversial question is whether state policies over the past decade have made any difference in raising student achievement.

*Quality Counts at 10: A Decade of Standards-Based Education* tracks state policy initiatives over the past decade in four core areas — standards, assessments, accountability, and efforts to improve teacher quality — based on 24 specific indicators.

The report documents a steady increase in the number of states embracing state content standards and tests linked to those standards between 1997 and 2005. Fewer states adopted all of the accountability and teacher-quality indicators tracked for the report during the same period.

To examine the relationship between standards-based education and gains on NAEP, the Editorial Projects in Education Research Center conducted a series of regression analyses for *Quality Counts 2006*. The independent variable was changes in the strength of states' standards-based policies between 1997 and 2005. The dependent variable was changes

in NAEP math achievement between 1996 and 2005, and in NAEP reading achievement between 1998 and 2005.

The center conducted separate analyses for math and reading in grades 4 and 8. To avoid biasing the results, the analyses controlled for states' initial NAEP performance at the start of the period, and for the initial strength of their standards-based policies.

One way in which the research center's analysis differs from other studies is that it considered the overall effect of states' standards-based policies — across the areas of standards, assessments, accountability, and efforts to improve teacher quality — and looked at which of those policy areas, relatively speaking, independently contributed the most to changes in achievement.

Initially, the analyses found a moderate positive relationship between states' overall embrace of standards-based education and gains in student math achievement. The researchers observed a slight negative relationship for reading. Further analyses exploring the relative contribution of the four major policy components revealed that the implementation of policies related to teacher quality related negatively to achievement gains in both reading and math.

In a second analysis, the researchers eliminated the teacher-quality policies from the overall measure of standards-based-reform implementation in order to focus specifically on the contributions of policies related to standards, assessments and accountability. Once teacher quality was taken out of the picture, the relationship between states' efforts to implement standards-based reforms and gains in student achievement became much stronger.

Improvement for math in grades 4 and 8 became statistically significant, while more modest, but positive, effects emerged for reading.

Preliminary analyses also found no relationship between state resource and equity indicators and student-achievement gains, after controlling for states' initial NAEP performance.

## **Such Exploratory Analyses Have Their Limits**

---

Gains on NAEP may fall short of measuring changes in student achievement at the state level, in part because NAEP tests are given only to a sample of students in each state, do not carry any stakes for individual students or schools, and may not completely match state content standards and priorities. States also vary in the specific policies they employ, when they put them into effect, and how consistently and coherently they carry them out. From a statistical perspective, the analysis also relies on a small sample, the 50 states. That provides another reason for caution in interpreting the results.

Despite those caveats, the time and energy put into standards-based education over the past decade suggest it's appropriate to examine the effects of state policies as a way to help inform state practices in the future. The results suggest that while standards-based education and its implementation are far from perfect, they can contribute to improvement in American schools.

As part of this 10-year retrospective on standards-based education, *Education Week* also invited five prominent policy observers to contribute their personal views to *Quality Counts* on what standards-based policies have accomplished so far, and what the next phase of improvement steps should be.

As is true every year, the 2006 report also tracks student achievement across the 50 states and the District of Columbia, and charts progress on states' education systems in four areas: standards and accountability, efforts to improve teacher quality, school climate, and school resources and the equity of school finance systems. States averaged a C+ across the graded categories this year.

For the first time, the Editorial Projects in Education Research Center also has produced detailed state-by-state reports on how states have performed on this year's indicators and the progress they've made over time. The state-highlights reports, which replace and expand on the state summaries that appeared in previous print editions of *Quality Counts*, can be viewed at [www.edweek.org/qc06/shr](http://www.edweek.org/qc06/shr).

*Quality Counts is produced with support from the Pew Center on the States.*

# THE NO CHILD LEFT BEHIND ACT OVERVIEW

---

President Bush signed the No Child Left Behind Act into law in January 2002. The law significantly strengthens the federal government's role in elementary and secondary education. Notable features of NCLB are its focus on improving the performance of poorly performing schools, and accountability and results aimed at raising academic achievement for all students. Other notable features include:

- an emphasis on closing the achievement gap
- requirements that states hire only “highly qualified” teachers
- the Reading First program designed to boost reading proficiency of children in K-3
- the focus on using scientifically based research to support new programs and practices
- the appropriation of significant new federal dollars to provide high-quality professional development to support teachers in the classroom

Testing is a vital part of standards-based reform. All states were required to implement reading and math assessments in grades 3-8, and once during grades 10-12 beginning in the 2005-06 school year. By school year 2007-08, all students must be tested in science at least once. Those new assessments must be aligned with state academic standards, must produce results that are comparable from year to year, must show whether students are meeting the state standards, and the results must be reported widely. Considerable federal funds are appropriated to make the reforms happen. Deadlines are set, and penalties are spelled out if the reforms do not happen.

For a more detailed review, consult:

- U.S. Department of Education's No Child Left Behind resource Web site at [www.ed.gov/nclb](http://www.ed.gov/nclb)
- Regulations that have relaxed some of the requirements for special education, rural education, and more at [www.idea.practices.org](http://www.idea.practices.org).

*The spread of freedom is spurring technological innovation and global competition at a pace never before seen. Billions of new competitors are challenging America's economic leadership — and if we are going to keep up — we're going to have to run faster.*

— Margaret Spellings

*U.S. Secretary of Education*

# ETS ON THE ISSUES

## It's Time to Act on High School Reform

By Kurt Landgraf, President and CEO, ETS

---

"The imperative for action is urgent."

The quote comes from Achieve, a bipartisan group created by governors and business leaders to support reform of America's public high schools. The insight couldn't be more timely, or more worth heeding.

By now, it's widely recognized that our high schools are, as Microsoft Chairman Bill Gates says, "obsolete." In recent months, a consensus has also emerged that now is the time to act. Those of us who are passionate about education reform have begun to do just that.

The National Governors Association's 2005 education summit, for instance, saw a number of excellent reform proposals, including Gates' call for an emphasis on "the 3 R's: *rigorous* curriculum, made *relevant* by excellent teachers, in a setting in which personal *relationships* create a welcoming learning environment."

The governors' summit also produced a five-point "Action Agenda" that can serve as a model for school systems throughout the country:

1. Restore value to the high school diploma by raising standards for all students, and tying graduation tests and requirements to the expectations of colleges and employers.
2. Redesign high schools to provide students with the high-level knowledge and skills they need and the educational options and support that are crucial to academic success.
3. Give students the excellent teachers and principals they need by improving teachers' and principals' knowledge and skills, and providing incentives for teachers to go and stay where they're needed most.
4. Set goals, measure progress, and hold high schools and colleges accountable to benchmarks, and provide timely, effective intervention in low-performing schools.
5. Streamline and improve education governance so that elementary, secondary and postsecondary schools work more collaboratively.

Collaboration between secondary and post-secondary schools is key. It will be the theme of ETS's own "College Readiness/College Success Summit" next month, which will bring together some of the leading groups involved in helping to ensure high school students are prepared for post-secondary success.

Now that high school reform is on the public policy agenda, we have an opportunity to make improvements that will last for generations — if we seize the moment.

At ETS, we're doing our part. We're listening to educators, parents and policymakers. We're learning from sound research. And we're leading the effort to achieve both informed public policy and informed educational practice.

# Putting Assessments to Work in the Classroom

*By Kurt Landgraf, President and CEO, ETS*

---

The public conversation about the quality of American education evolved and intensified over the past two decades, and now real reform is being achieved. Some may want to continue arguing about standards, accountability and testing, but the real battle is over. National public policy has been set. We will have standards. We will have accountability. And assessments are going to be an important element in the equation.

ETS is ready to join with teachers and school leaders in taking the next essential step, to put assessments to use where they matter most: in the classroom. We must make much better use of tests and test results so that we can determine who's learning, who's not, and how classroom practice and professional development can be adjusted to ensure that no student is, in fact, left behind.

Unlike tests used for high-stakes decisions such as whether a student graduates, classroom or "interim," assessments measure what students are learning during the course of the school year.

Do they work? Absolutely. As San Francisco's Bay Area School Reform Collaborative found, effective use of data produced by assessments of student achievement is instrumental in narrowing the achievement gap among students of widely different backgrounds. Schools that are successful in reducing

the achievement gap use assessments more frequently, and are more likely to use the data produced by those assessments, than do less successful schools. That may not be a surprising finding, but it certainly is an important one.

Too often, teacher education students as well as practicing teachers are shortchanged when it comes to learning how to make effective use of assessments.

At ETS, we are hard at work creating tools to bridge this unfortunate and avoidable gap. For example, our Understanding Standards-Based Assessments series of professional development workshops for teachers demonstrates how to monitor each student's level of achievement, and how to collect and organize the information and feedback necessary to help shape effective instruction.

Timely, targeted assessments provide educators with the information they need to improve teaching and learning. High-quality assessments, properly used, help identify student-learning problems, their causes and workable solutions.

At ETS, we're listening to educators, parents and policymakers. We're learning from sound research. And we're leading the effort to achieve both informed public policy and informed educational practice.



# Teacher Quality: Defining Good Teaching

## **Attract the Most Qualified and Retain the Highest Quality**

---

The formula for achieving a highly qualified teaching force is no secret. Maintain a rigorous standard for entry; give early support for beginning teachers, including training and mentors; pay for higher salaries and better working conditions; and make retaining the best teachers a priority. Start with raising the standards for teacher preparation and licensing. Only half of America's teacher education schools meet national accreditation standards. All should. First-year teachers have a 40 percent turnover rate, but that can be cut in half with high-quality support programs. Better pay attracts skilled candidates and keeps quality teachers in the profession. Evaluate, support and retain these high-quality teachers.

## **Effective Professional Development Aligns Teachers with Standards**

---

Retaining quality teachers requires professional development throughout their careers. Such development, if effective and ongoing, can align teachers' classroom methods with high standards and the demanding curricula to which they are accountable. Routinely evaluating teachers in practice, their teaching skills, and content knowledge, must be done — but this alone is not enough. These steps must lead to programs that develop high-quality teachers current in their fields and sharpened in their skills. These teachers will benefit from scientifically based research, and their students will gain in learning. Without this professional support, teachers may not remain in the profession long enough to make a difference in students' lives.

## **Why Not a Highly Qualified Teacher in Every Classroom?**

---

Attract highly skilled candidates to the profession, then steer them to where they are needed the most — to underserved and understaffed locations, subjects and specialties. Where the needs are the clearest and the challenges the most daunting, the rewards should be the greatest. Quality teaching must get in the way of the pervasive inequality that blocks a healthy democracy. Here, the highly qualified teacher will be one who knows his or her subject, knows how to teach it, and actually shows that he or she is able to teach it. The highly qualified teachers will be a professional who “combines passion and art along with academic prowess.”\*

## Teacher Quality

*If we want better teachers, we have to treat our teachers better.*

— Kurt Landgraf, President and CEO, ETS, “Solving the Teacher Shortage”

---

\**Teaching at Risk: A Call to Action*, the Teaching Commission report, January 2004.

# WHERE WE STAND ON TEACHER QUALITY: AN ISSUE PAPER FROM ETS

---

## Introduction

---

ETS, the world's largest private educational testing and measurement organization, has more than 50 years of experience in the area of teacher quality. We develop teacher, administrator and paraprofessional assessments; produce related professional development products and services; and conduct program and policy research on education personnel and practices.<sup>1</sup> In these roles, we are often asked for advice and counsel about improving the quality of teachers. In order to share our understandings with a wider audience, ETS has decided to publish them in a series of issue papers on improving the quality of the teacher workforce in the United States.

This first paper in the series, entitled "Where We Stand on Teacher Quality," addresses aspects of teacher quality that we think are fundamental. Some of our conclusions are based on compelling research; others, in the absence of hard data, on our own and others' experience and on the best available professional judgment. The issues included here are not comprehensive. We have selected those aspects of teacher quality that ETS, because of its experience and expertise, has a particular standing to address.

More specific papers in coming months will focus on particular topics we think warrant further examination. They will address, in general, closing the achievement gap and the related teacher-quality gap.

## What It's All About

---

Good teachers produce good students.<sup>2</sup> A teacher's most important task is contributing to and enhancing the learning and achievement of his or her students.

What are the characteristics of teachers who consistently improve the achievement of their students? A large segment of the education research community is engaged in attempting to answer that question. Promising progress is being made in creating data systems to identify the attributes and strategies of teachers whose students, over periods of several years, have achieved unusually large gains in the subjects taught.<sup>3</sup> ETS is investigating the challenges of aligning teacher data with student data; linking teacher performance to student learning; and determining the accuracy, fairness and appropriate use of those methods and measures. We are also conducting a study to investigate the relationship between teachers' performance on licensure tests and student achievement.

The remainder of this paper describes where ETS stands on teacher quality and examines that important topic within the context of a teacher's progression along the continuum of teaching — from teacher preparation through ongoing professional development and performance evaluation. It includes recommendations for:

- defining teacher quality
- raising standards for those entering the profession
- supporting teachers throughout their careers

---

1. More specifically, *The Praxis Series: Professional Assessments for Beginning Teachers*®, customized state teacher licensure tests, the *Parapro* Assessment for paraprofessionals, assessments of accomplished teaching for the National Board for Professional Teaching Standards, the School Leaders Licensure Assessment, and the School Superintendent Assessment. ETS also produces the *Pathwise*® Series of professional development materials, workshops, training sessions, software, and minicourses for teachers; the *Pathwise* Framework Induction Program, Observation Program, Portfolio Program, and Leader Academy; as well as the *Pathwise* Teacher Evaluation System, the *Pathwise* School Leadership Series, and the Pathwise Understanding Assessment Series. See: [www.ets.org/pathwise/index.html](http://www.ets.org/pathwise/index.html).

2. Jennifer King Rice, *Teacher Quality: Understanding the Effectiveness of Teacher Attributes* (Washington, DC: Economic Policy Institute, 2003).

3. Charles T. Clotfelter, Helen F. Ladd, and Jacob Vigdor, "Do School Accountability Systems Make it More Difficult for Low-Performing Schools to Attract and Retain High-Quality Teachers?" *Journal of Policy Analysis and Management* vol. 23, no. 2 (2003), pp. 251-257.

## Defining Teacher Quality

---

We know certain things about teacher quality. A correlation exists between teacher verbal ability and student achievement. Teachers who majored in the subject they teach are better teachers of that subject than those who did not. Pedagogy, particularly content-based pedagogy, has a positive impact on student achievement, and teachers with considerable experience are likely to make a greater contribution to student learning than teachers with few years of teaching experience.<sup>4</sup>

While there is some evidence about the qualifications of those who enter the profession, there is a need for more information and research on the quality of teachers' classroom practices, on the impact of those practices over time, and on how professional development programs support improved teaching and learning.<sup>5</sup> Any definition of teacher quality, of course, must be made within the context of the NCLB and its "highly qualified teacher" provisions.

**Knowing Content.** States are attempting to comply with the NCLB requirement that all teachers of core subjects be highly qualified by the school year 2005-06. The definition of a highly qualified teacher contained in NCLB is based on the above-noted, widely accepted conclusion that good teachers must know the subjects they teach.

To be highly qualified, according to NCLB, teachers must meet three general requirements: have a bachelor's degree, be licensed or certified by

the state, and demonstrate *subject matter competence* in each academic subject they teach. The definition differs depending on the level of students being taught and whether a teacher is new or already teaching. *New elementary* teachers must demonstrate their knowledge and teaching skills by passing a rigorous state test in reading, writing, math and other areas of the elementary curriculum. *New middle/secondary* teachers may demonstrate competency in each academic subject they will teach by having a major or equivalent coursework, a graduate degree or an advanced credential — or, alternatively, by passing a rigorous state academic subject test — in each subject to be taught. *Current* teachers at any level may demonstrate competence as noted above or by meeting a "High Objective Uniform State Standard of Evaluation," also known as a HOUSSE.

ETS and others<sup>6</sup> are concerned about the rigor of the HOUSSE provisions and about reports that some states may be using them as ways to circumvent teacher accountability provisions intended by the law. Veteran teachers with three or more years' experience should have an opportunity to demonstrate that they have achieved a high level of proficiency in their content area and also in the skills and art of actual classroom practice.

---

4. Daniel Fallon, "Case Study of a Paradigm Shift: The Value of Focusing on Instruction" (speech given at the fall Steering Committee meeting of the Education Commission of the States, Richmond, VA, November 12, 2003, [www.ecs.org/clearinghouse/49/00/4900.htm](http://www.ecs.org/clearinghouse/49/00/4900.htm)).

5. Harold Wenglinsky, *Teaching the Teachers: Different Settings, Different Results* (Princeton, NJ: ETS, Policy Information Center, 2000); Barbara A. Bruschi and Richard J. Coley, *How Teachers Compare: The Prose, Document, and Quantitative Skills of America's Teachers* (Princeton, NJ: ETS, Policy Information Center, 1999); Drew H. Gitomer, Andrew S. Latham, and Richard Ziomek, *The Academic Quality of Prospective Teachers: The Impact of Admissions and Licensure Testing* (Princeton, NJ: ETS, Teaching and Learning Division, 1999).

6. For example, The Education Trust and The Teaching Commission.

Thus ETS recommends that states:

- raise the bar for content rigor by reviewing and increasing passing scores on teacher licensure content examinations
- grant a content “pass” to all veteran teachers who reached the new passing score when initially assessed
- institute an ongoing series of rigorous, uniform performance assessments, to be administered to veteran teachers by trained reviewers following a standard observation protocol
- utilize teacher performance data in concert with student achievement data as indicators of teacher quality

***Knowing How, and Being Able to Teach.***

ETS has elected not to participate in the debate over the importance of content knowledge versus teaching skills (or pedagogy) in defining good teaching. We think that it creates an inappropriate dichotomy and offers a false choice. Knowing one’s subject, knowing how to teach it, and actually being able to teach it are all critical.

The public holds this view as well. According to *Quality, Affordability, and Access: Americans Speak on Higher Education*, a public opinion poll conducted by pollsters Hart and Teeter, one-third or

more of the public cites having the skills to design inspiring learning experiences and having a caring attitude toward students as the two most important qualities for a good K-12 teacher.<sup>7</sup>

ETS thinks that competent teachers should have the following knowledge and skills:

1. Basic academic reading, writing and math<sup>8</sup>
2. Thorough knowledge of the content of each subject taught, appropriate to the levels of their students<sup>9</sup>
3. Both generic and content-specific knowledge in areas such as child development, classroom management, motivating children to learn, interpreting and using assessment data, individualizing instruction, aligning content to the state’s standards, developing appropriate instructional materials, and working with children with disabilities or from other cultures<sup>10</sup>
4. Hands-on ability and skill to use the above types of knowledge to engage students successfully in learning and mastery<sup>11</sup>

Increased attention to the importance of non-cognitive measures is under way. The importance of persistence, enthusiasm, communication skills, and the belief that all students can learn are important factors that ETS and others are studying.

7. Peter Hart and Robert Teeter, *Quality, Affordability, and Access: Americans Speak on Higher Education* (Washington, DC, 2003). Opinion poll commissioned by ETS.

8. As demonstrated through test scores, see: Ronald Ferguson, “Paying for Public Education: New Evidence on How and Why Money Matters,” *Harvard Journal of Legislation* vol. 28 (Summer 1991): 465-498; Brian Rowan, Fang-Shen Chiang, and Robert J. Miller, “Using Research on Employees’ Performance to Study the Effects of Teachers on Students’ Achievement,” *Sociology of Education* vol. 70 (1997): 256-284; and David H. Monk, “Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement,” *Economics of Education Review* vol. 13 (1994), pp. 125-145.

9. As demonstrated through degree or coursework in the subject taught – especially in mathematics – see: Dan D. Goldhaber and Dominic J. Brewer, “Why Don’t Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity,” *The Journal of Human Resources* vol. 32 (1997), pp. 505-23; Brian Rowan, Fang-Shen Chiang, and Robert J. Miller, 1997; and David H. Monk and Jennifer A. King, “Multilevel Teacher Resource Effects in Pupil Performance in Secondary Mathematics and Science: The Case of Teacher Subject-Matter Preparation,” in *Choices and Consequences: Contemporary Policy Issues in Education*, ed. Ronald G. Ehrenberg, pp. 29-58 (Ithaca, NY: ILR Press, 1994).

10. As demonstrated through appropriate teacher certification, see: David H. Monk, 1994, and Linda Darling-Hammond and Peter Youngs, “Defining ‘Highly Qualified Teachers’: What Does ‘Scientifically Based Research’ Tell Us?” *Educational Researcher* vol. 31 (2002), pp. 13-25.

11. On classroom practice as an indicator of teacher quality, see: Harold Wenglinsky, *How Teaching Matters: Bringing the Classroom Back Into Discussions of Teacher Quality* (Princeton, NJ: ETS, Policy Information Center, 2000)

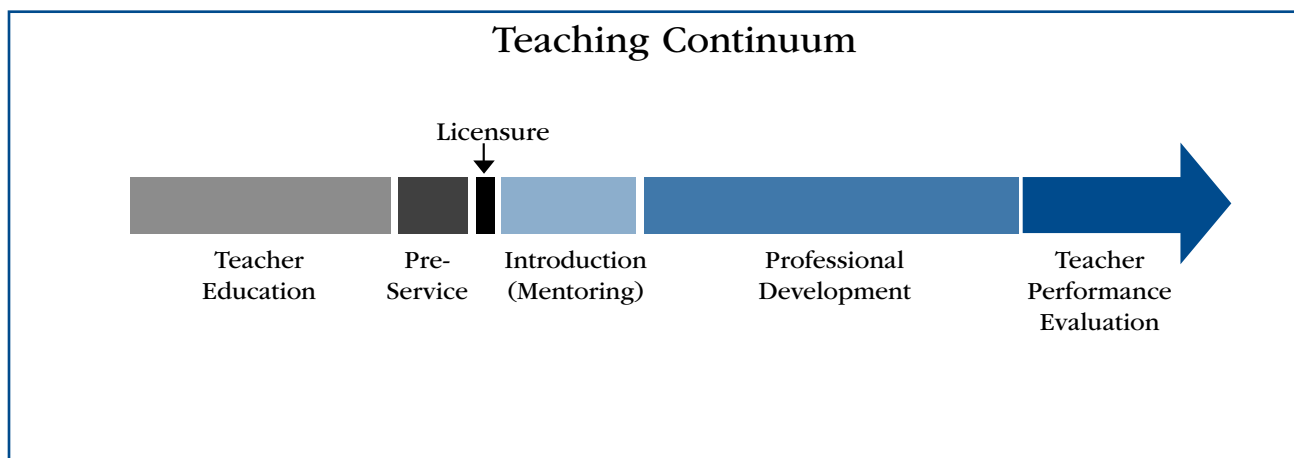
## The Teaching Progression & Continuum

Teachers do not enter the classroom as finished products. Most teachers improve and grow over time.<sup>12</sup> When they enter the classroom, new teachers do not possess all the knowledge and skills they need to become highly effective, but with experience, practice, assistance and training, novices can become better teachers.

Often overlooked is that NCLB defines another type of teacher, an “exemplary” teacher. That is not a beginning teacher, but an experienced teacher that the law describes as one who meets the “highly qualified” requirements and is recommended as exemplary by administrators and others; mentors other teachers, helping them improve their instructional strategies and skills; develops curricula; and offers other forms of professional development.

That is the ultimate, highest quality teacher sought for every classroom — and that ETS is striving to help develop through our work for the National Board for Professional Teaching Standards (NBPTS).

ETS develops assessments used by the National Board to measure the performance of certification candidates against high and rigorous standards. We see the profession of teaching as a continuum, from pre-service preparation to initial licensure, to hiring and induction, continued practice and professional development through mastery and advanced certification. The graphic below offers a time-line perspective illustrating the transitions in the developing career of a teacher. Particularly striking is the brief period from preparation (Teacher Education) to hiring (Licensure), and the lengthy expanse of on-the-job classroom teaching (Professional Development and Teacher Performance Evaluation). The pre-employment phase consumes only about one-tenth of a 40-year, lifetime teaching career. It is during the time spent teaching in the classroom that teachers, given proper support, evaluation, feedback and targeted professional development, can evolve and mature into accomplished, exemplary teachers.



12. National Commission on Teaching and America's Future, *No Dream Denied: A Pledge to America's Children*, Summary Report (Washington, DC: NCTAF, 2003), 20.

A critical factor in developmental progress is *the alignment* of the state student content standards throughout the continuum. It is essential for both equity and efficacy in helping students meet the standards on which they will be tested — and on which, increasingly in the future, teachers' performances will be based. Alignment of state standards should be evidenced in pre-service preparation, licensure exams, curricula, related professional development efforts and teacher performance evaluations. Just as students must be exposed to curricula linked to the standards upon which they are assessed, teachers must be prepared to teach such curricula before their effectiveness is evaluated.

ETS is committed to enhancing the alignment of our tests and professional development products for teachers and school leaders with state student content standards.

For our *Praxis*<sup>™</sup> licensure series, we develop specifications for each test that are aligned with the knowledge requirements of the entry-level teacher's job and with the appropriate teaching curriculum standards established by the nation's organizations for professional teaching disciplines.<sup>13</sup> ETS also works with each client state to identify those subject tests that best align with the state's content standards for entry-level teachers teaching subject areas and/or grade levels.

*We should connect teacher professional development with teacher preparation standards, student standards, curriculum and assessments to achieve an aligned system of preparing and supporting new and in-service teachers.*<sup>14</sup>

— Kurt Landgraf, President and CEO, ETS

## Teacher Preparation

---

**Accountability.** Title II of the Higher Education Act, currently being reauthorized by Congress, includes accountability provisions for the nation's teacher education programs. A key requirement is the annual reporting of the pass rates of each institution's graduates on state teacher licensure exams. ETS provides the institutional pass-rate information in states using the *Praxis* assessment series.

ETS thinks Title II should be redirected to focus on strengthening teacher education programs<sup>15</sup> so as to enhance the quality of program graduates. ETS President Kurt Landgraf, in congressional testimony, said, "It is not clear that pass-rate-based standards lead to improvement of teacher education programs or to the quality of their graduates."<sup>16</sup>

ETS also supports the recommendation of the National Research Council's recent report against using pass rates on teacher licensing tests as the sole basis for determining the performance of teacher education programs.<sup>17</sup>

We think pass-rate data are important but additional information is needed, including data on candidates' actual scaled scores, to provide more useful information about the quality of preparation programs. Yes/No pass-rate data do not indicate the range or the distribution of candidate scores, providing information of limited use as a result.

13. For example, *Praxis* science exams are aligned with the standards of the National Science Teachers Association.

14. Kurt M. Landgraf, testimony on teacher preparation and accountability at hearing on "Training Tomorrow's Teachers – Providing a Quality Post-secondary Education," before the Twenty-First Century Competitiveness Subcommittee of the U.S. House Education and the Workforce Committee, October 9, 2002.

15. See Ronald G. Ehrenberg and Dominic J. Brewer, "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond," *Economics of Education Review* vol. 13, no. 1 (1994), pp. 1-17; and Harold Wenglinsky, *Teaching the Teachers: Different Settings, Different Results*.

16. Kurt M. Landgraf, testimony on teacher preparation and accountability.

17. National Research Council, *Testing Teacher Candidates: The Role of Licensure Tests in Improving Teacher Quality* (Washington, DC: National Academy Press, 2001).

Further, different states set different passing scores on the same licensing tests, making cross-state comparisons impossible. Thus, using pass-rate data in concert with scaled scores and other output measures about the program may provide reason to remediate or celebrate.

Even with additional output measures, however, it is important to know the characteristics of the teacher candidates. Prospective teachers vary according to their literacy skills, content knowledge, and motivation to succeed — qualities not equally distributed among candidates at various institutions. Thus, it is difficult to ascertain whether high test scores are due to significant student growth resulting from an institution's rigorous preparation program or to the pre-existing skills and drive of the students who chose to attend. It is therefore imperative that multiple indicators be used to measure the effectiveness of teacher education programs.

**Clinical experience.** ETS supports the inclusion of clinical teaching experience over a sustained time period as a part of teacher preparation and as a condition of program approval and licensure. Recent data show that a few remaining states do not require clinical experience during teacher preparation<sup>18</sup>; we recommend that they do so.

The United States was the only one of the seven high-performing countries on TIMSS 99 that did not have a high-stakes evaluation of practical experience as part of its teacher education requirements.<sup>19</sup> We think such experience helps develop teaching competence, confidence and retention on the job,

and that continued observation, monitoring and feedback are also critical. The lack of clinical skills and classroom experience has been found to be linked to teacher burnout and attrition among new teachers.<sup>20</sup>

We are particularly cognizant of the Carnegie Corporation's exciting and promising Teachers for A New Era (TNE) teacher-education reform initiative. TNE does not simply call for a practice teaching component; it redefines teaching itself as an academically taught clinical practice profession, requiring close cooperation between colleges of education and actual practicing schools. Eleven outstanding teacher preparation institutions are beginning to transform their education programs into schools of modern clinical practice, providing their graduates with clinical residency programs similar to those in the field of medicine.

**Recognizing Excellence.** ETS is working on several fronts to help districts, states, education leaders and policymakers raise the standards for those entering the profession. The new Recognition of Excellence program, similar to a college honors diploma, recognizes and encourages exceptional individual performance on select *Praxis II*<sup>®</sup> tests. Candidates who earn high scores (in the top 15 percent of test takers) on any of 11 *Praxis II* tests will receive a certificate from ETS, and the award will be noted on all *Praxis* score reports.

18. Ronald A. Skinner and Lisa N. Staresina, "State of the States – Efforts to Improve Teacher Quality," *Education Week, Quality Counts 2004: Count Me In* vol. 23, no. 17 (2004), p. 110.

19. Aubrey H. Wang and others, *Preparing Teachers Around the World* (Princeton, NJ: ETS, Policy Information Center, 2003), 7, 21-23, and 44. [TIMSS 99 refers to the repeat, in 1999, of the 1995 Trends in International Mathematics and Science Study, for students in the eighth grade.]

20. National Commission on Teaching and America's Future, *No Dream Denied*, 20; and Erling E. Boe and Sharon A. Bobbitt, "Why Didst Thou Go? Predictors of Retention, Transfer and Attrition of Special and General Education Teachers from a National Perspective," *Journal of Special Education* vol. 30 (4), pp. 390-411.

## Teacher Licensure

---

**What It Means.** Licensure is an important element to assure quality in the teaching workforce of each state. The purpose of teacher licensure, however, is often misunderstood. Licensure is used to identify those teacher candidates with the knowledge and skills deemed important for a beginning practitioner. Initial licensure represents a beginning for the novice. The newly certified candidate has earned access to a profession that, like others, is characterized by an upward progression of knowledge and skill, achieved over time through additional learning, practice, experience and maturity. Licensure establishes a floor; it does not ensure excellent practice. That determination can only occur in the context of teaching.

Typically, states offer an initial provisional license at the time of hiring. After serving successfully for several years on the job, the teacher may then apply for a permanent license, which is subject to renewal.

**Key Components.** ETS thinks that whatever the route, responsible licensing decisions include three key components: education, experience and examination.

The role of the teacher in advancing learning is so important that education, an examination or experience alone will not provide a sufficient basis to make an appropriate decision regarding licensure. All three are critically important.

**Licensure Tests.** Tests are usually one of several criteria necessary for licensure. Licensure tests are intended to distinguish between those who have a desired level of competence to begin practice from those who do not. The tests are knowledge measures — knowledge of basic skills, of subject matter and of how to teach. They do not, however, measure

competence in the classroom, nor do they indicate how well a teacher communicates with students or engages students' interest in learning. The *Praxis III*<sup>®</sup> performance-based assessment does, however, offer an observation-based evaluation of beginning teachers' classroom performance.

Historically, licensure tests have not been designed to predict performance. For example, the bar examination does not predict how good a lawyer a student will become or the number of cases that person will win. Similarly, teacher licensure tests have traditionally not been held to the standard of predicting outcomes — but the changing environment in K-16 accountability has produced increased interest in holding them to this higher standard.

In developing *The Praxis Series*<sup>™</sup> exams, ETS has addressed how and what teachers should learn in order to gain entry to the profession. We have analyzed the 49 sets of state student standards in every content area in which they are available. We have combined that with national disciplinary organization standards and based the content of the teacher licensure tests on the many areas in which all standards agree. We have consulted with teachers and subject-matter experts across the country in developing the *Praxis* assessments.

**Helping Candidates Succeed.** To help teacher candidates prepare for the *Praxis* assessments, ETS has published learning guides for 27 of the subjects we test. Each guide presents an architectural diagram of the critical foundations of the content domain of each test. Our new *Praxis*™ Diagnostic Preparation Program provides detailed, customized feedback about candidates' performances so they can better understand their strengths and weaknesses and can focus their test preparation efforts accordingly.

**Alternative Routes.** ETS supports the concept of alternative routes to teaching, including reducing unnecessary barriers and expediting the process to encourage the entry of new, talented candidates into the field. While the relative emphasis placed on each of the three factors noted above may change in order to open the door to prospective teachers, regardless of what route a teacher follows, all three components are needed. We support public policies that require graduates of alternative programs to meet the same accountability requirements as their traditionally prepared peers.

**Licensure Reform.** ETS thinks that through its impact on entry to the profession, licensure offers great potential to enhance teacher quality across the country. However, licensure reform is necessary. Given that no national licensure standard exists, and state practices and policies vary considerably, ETS encourages cross-state collaboration to expedite progress.

**Raising the Bar.** States should re-evaluate existing teacher licensure programs and raise the standards for those entering the profession. Specifically, states should work toward a common standard for subject-area tests and set cutoff scores at a level that requires teaching candidates to demonstrate mastery reflecting college-level study.<sup>21</sup>

**Uniformity, Comparability and Portability.**

*The Praxis Series* assessments are a national program, with the same tests provided to all states that select them. Only the variation in passing scores precludes comparability. Portability of scores, and thus candidate mobility and reciprocity across states, would be greatly enhanced with uniform passing scores. ETS is committed to working with states to reduce differences in passing scores on *Praxis* tests across states. States can begin by reviewing standards required for demonstration of subject-matter knowledge, including the passing scores required on these assessments. We also commit to working with our client states to develop more-uniform methods of evaluating clinical experience.

We are working with an exciting collaboration, the Mid-Atlantic Regional Teachers Project, to develop full regional licensure reciprocity, new-teacher mentoring programs, common regional standards for alternative certification, regional pension portability, and a new regional designation of "meritorious new teachers."

**Streamlining the Process.** We agree with calls to streamline the cumbersome bureaucracy that often surrounds teacher licensure in order to make the profession more attractive to a wide range of qualified candidates.<sup>22</sup> States should consider reducing the number of licenses and working together toward more uniformity across states. ETS is collaborating with Teach for America, offering *Praxis* tests at convenient times to help accelerate Teach for America candidates' entry to the classroom. We are also offering flexibility to candidates of The New Teacher Project in the *Praxis* registration process.

21. The Teaching Commission, *Teaching at Risk: A Call to Action* (New York: The Teaching Commission, 2004), pp. 40-1.

22. *Ibid.*

## Induction and Mentoring

---

ETS urges all states to establish beginning-teacher induction programs that provide mentoring and other support during the crucial first years of teaching. Research shows that teachers without induction support leave the profession at rates almost 70 percent higher than those who receive it.<sup>23</sup> With about one-third of new teachers leaving the classroom within three years and nearly one-half within five years,<sup>24</sup> not providing induction is an irresponsible decision. Yet, only 15 states both require and finance mentoring programs for all novice teachers<sup>25</sup> — a dismal record, especially since significant federal funds are available for that purpose. Others have echoed that conclusion, recommending a mentor for all new teachers throughout their first year. As asserted by The Teaching Commission: “Nobody should simply be thrown into a classroom without the support needed to be successful.”<sup>26</sup>

In his October 2002 testimony before Congress, ETS President and CEO Kurt Landgraf highlighted the California Formative Assessment and Support System for Teachers (CFASST), an assessment process designed to help beginning teachers grow in their profession. Modeled after ETS’s Framework for Teaching, CFASST supports beginning teachers in their development through a dynamic, ongoing process — planning and teaching lessons, reflecting on the results, and then making informed instructional changes. Results of a recent study commissioned by the California Commission on Teacher Credentialing examined the impact of its Beginning Teacher Support and Assessment (BTSA)/CFASST on the teaching effectiveness of beginning teachers and on the learning of the students of these teachers.

Overall, the findings showed a positive impact of BTSA/CFASST program on teachers and students and confirmed the model of how BTSA/CFASST works: the support of an experienced teacher, the curriculum of CFASST events, and the improved practices, in turn, lead to improved student learning.<sup>27</sup> ETS is also working with Ohio, Michigan, Iowa, Virginia and Mississippi on teacher induction.

Essential to good mentoring programs is a cadre of highly skilled mentors who are able to quickly diagnose the needs of struggling teachers and constructively relate to and support them. ETS’s Framework Observation Program gives mentors, administrators and coaches crucial training and tools for conducting in-depth observations based on evidence, not opinion, and for providing consistent feedback. It also offers a structured process for enhancing practice through planning, reflection and analysis.

## Ongoing Professional Development

---

Continuous professional development is critical to developing and maintaining high-quality teachers. Data show that without highly skilled support, even those with high qualifications will not remain in the profession long enough to make a difference in children’s lives. The Teaching Commission urges principals to ensure that all teachers have the chance to improve their classroom instruction by receiving ongoing training aimed at professional growth and better student outcomes. Principals should “be held responsible for ensuring that all teachers benefit from scientifically based professional development opportunities that focus squarely on assessing and improving instructional practices and thereby raising student achievement.”<sup>28</sup>

---

23. U.S. Department of Education, National Center for Education Statistics, *Progress through the Teacher Pipeline: 1992-93 College Graduates and Elementary/Secondary School Teaching as of 1997* (Washington, DC, 2000). NCES 2000-152.

24. National Commission on Teaching and America’s Future, *No Dream Denied*, 24.

25. Ronald A. Skinner and Lisa N. Staresina, “State of the States – Efforts to Improve Teacher Quality,” 112.

26. The Teaching Commission, *Teaching at Risk: A Call to Action*, 48-9.

27. Marnie Thompson and Pamela Paek, “Implementation and Impact of BTSA/CFASST” (paper presented at the National Induction Conference, Sacramento, CA, March 10-12, 2004. Revised March 30, 2004).

28. The Teaching Commission, *Teaching at Risk: A Call to Action*, 47.

The development of each of the *Pathwise*<sup>®</sup> products for professional development, designed to improve teacher and school leader performance, are grounded in what research studies define as “best practice.” More importantly, however, ETS is committed to conducting its own rigorous long-term, scientifically based research, designed to show either that those products do generate gains in student learning, or that we must modify the programs to bring about such gains.

## **Teacher Performance Evaluation**

---

Evaluations of teachers’ performance in the classroom occur at many points on the teaching continuum, at various times throughout a school year and for a variety of purposes. Those can include relatively low-stakes evaluations, such as ones that provide remediation to a teacher candidate, determine assistance required by a novice, or identify the effects of a professional development program. Alternatively, they may be evaluations for high-stakes decisions such as licensure, hiring, performance pay, promotion or tenure.

ETS thinks strongly that the teaching skills and content knowledge of teachers should be routinely observed in the classroom and evaluated throughout their careers. Unfortunately, performance evaluation is frequently a missing element of teacher development

planning and, as noted earlier, when used to assist fledgling candidates, can mean the difference between leaving and staying.

ETS has developed the strategies, methods and techniques for appropriate and accurate direct observation of teacher performance. These are reflected in the assessments of the National Board for Professional Teaching Standards<sup>™</sup> and in the *Praxis III*<sup>®</sup> Performance Assessments for second-tier licensure. They are also incorporated in our *Pathwise*<sup>®</sup> Framework Observation program.

We recognize that additional research is needed in order to utilize the emerging data on student achievement and to relate it to teacher performance evaluations. ETS has developed a long-term policy research agenda to investigate this and related issues, and we are committed to studying the relationship between good teacher performance and improved student achievement. As the national leader in the field of assessment, we are leading the way in improving performance assessment for teachers and using student achievement data in the process.

We urge that high-quality performance evaluations be required as a part of licensure. We also support the use of individual teacher evaluations for performance pay determinations.

## Where We Stand on Teacher Quality

---

ETS stands ready to work with policymakers and practitioners to improve teacher quality and student achievement. From our perspective, strong content knowledge and knowing how to teach are both essential qualifications that beginning teachers must have to enter the classroom. We think that teachers who meet high qualifications for entry will grow and progress in their classroom teaching practice. We recommend that states work together to achieve more commonality and comparability in qualifications for those entering and staying in this important profession. More specifically, we urge public policymakers to adopt policies on three critical issues:

### Defining Teacher Quality

- Good teachers produce good students. That is the primary purpose of teaching and the basis upon which teachers should be evaluated.
- ETS eschews the debate over content versus teaching skills in defining good teaching. It creates an inappropriate dichotomy, a false choice.
- Good teaching requires four types of knowledge and skills: basic academic skills, thorough content knowledge of each subject to be taught, knowledge of both generic and content-specific pedagogy, and hands-on teaching skills.

### Raising Standards for Entry

- States should re-evaluate their teacher licensure programs and begin raising their entry standards, including the passing scores required on licensure exams.
- ETS pledges to work with states on efforts to reduce differences in passing scores on *Praxis*<sup>™</sup> tests across states.
- We support the inclusion of sustained clinical teaching experience in teacher preparation and as a condition of program approval and teacher licensure.
- Teacher candidate pass rate data are not enough; candidates' scaled test scores will provide additional, useful information about the quality of their teacher preparation.

### Supporting Teachers

- ETS urges all states to establish induction programs for beginning teachers, providing mentoring and support during the first years of teaching.
- Continuous high-quality professional development is critical to developing and maintaining high-quality teachers.
- Greater emphasis should be placed on observing and evaluating teachers' teaching skills and content knowledge in their actual classrooms throughout their careers.

# USING STUDENT PROGRESS TO EVALUATE TEACHERS: A PRIMER ON VALUE-ADDED MODELS

*A Policy Information Perspective by Henry Braun*

---

## Introduction

---

The most recent reauthorization of the Elementary and Secondary Education Act, the No Child Left Behind Act (NCLB), has been much more successful than its 1994 predecessor in galvanizing states into action. Undoubtedly, the main reason is the loss in federal aid that states would incur should they fail to comply with NCLB mandates — principally, those relating to schools and teachers. School accountability has a strong empirical component: primarily, a test-score-based criterion of continuous improvement, termed “adequate yearly progress” (AYP).

NCLB also requires states to ensure that highly qualified teachers are in every classroom, with “highly qualified” defined in terms of traditional criteria such as academic training and fully meeting the state’s licensure requirements. Focusing attention on teacher quality has been widely welcomed.<sup>1</sup> Interestingly, in this respect, some states have taken the lead by seeking an empirical basis for evaluating teachers, one that draws on evidence of their students’ academic growth.<sup>2</sup> Indeed, so the argument goes, if good teaching is critical to student learning, then can’t student learning (or its absence) tell us something about the quality of the teaching they have received? Although the logic seems unassailable, it is far from straightforward to devise a practical system that embodies this reasoning.

Over the past decade or so, a number of attempts to establish a quantitative basis for teacher evaluation have been proposed and implemented. They are usually referred to by the generic term

“value-added models,” abbreviated “VAMs.” Essentially, VAMs combine statistically adjusted test score gains achieved by a teacher’s students. Teachers are then compared to other teachers in the district based on these adjusted aggregate gains. Various VAMs differ in the number of years of data they employ, the kinds of adjustments they make, how they handle missing data, and so on.

There is a marked contrast between the enthusiasm of those who accept the claims made about VAMs and would like to use VAMs, on the one hand, and, on the other, the reservations expressed by those who have studied their technical merits. This disjuncture is cause for concern. Because VAMs rely on complex statistical procedures, it is likely that policy-makers, education officials, teachers and other stakeholders could all benefit from an understandable guide to the issues raised by the use of VAMs for teacher evaluation. (Although there is also considerable interest in using VAMs for school accountability, we will not address that topic here.<sup>3</sup>)

This report is designed to serve as such a guide, reviewing the strengths and weaknesses of VAMs without getting bogged down in methodological matters. It is organized in a Q&A format and draws on recent technical publications, as well as the general statistical literature.<sup>4</sup> The intent is to assist interested parties in their deliberations about improving teacher evaluation and to promote the responsible use of a powerful statistical tool.

---

1. See for example K. M. Landgraf, “The Importance of Highly Qualified Teachers in Raising Academic Achievement” (Testimony before the Committee on Education and the Workforce, U. S. House of Representatives, April 21, 2004.)

2. Such evaluations may be used to identify teachers in need of professional development, for administrative purposes (e.g., rewards and sanctions), or both.

3. There are both similarities and differences in the use of VAMs for school and teacher accountability.

4. This report draws heavily from D. F. McCaffrey et al., *Evaluating Value-Added Models for Teacher Accountability*, Santa Monica, CA: RAND Corporation, 2003. The most relevant parts of the statistical literature deal with drawing causal inferences from different kinds of studies. The classic reference is W. R. Shadish, T. Cook, and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston, MA: Houghton Mifflin Company, 2002.

# Questions About Measuring Value-Added Modeling

---

## 1. Why Is There Such Interest in Value-Added Modeling?

In almost all school districts, teacher evaluation is a notoriously subjective exercise that is rarely linked directly to student achievement. Developers of VAMs argue that their analysis of the changes in student test scores from one year to the next enables them to isolate objectively the contributions of teachers and schools to student learning. If their claims are correct, then we have at hand a wonderful tool for both teacher professional development and teacher evaluation.

One attraction of VAMs is that this approach to accountability differs in a critical way from the adequate yearly progress (AYP) provisions of the NCLB Act. To evaluate AYP, a school must compute for all students in a grade, as well as for various subgroups, the proportions meeting a fixed standard, and then compare those proportions with those obtained in the previous year. Observers have pointed out the problems arising from making AYP judgments about schools or teachers on the basis of an absolute standard.<sup>5</sup> The issue, simply, is that students entering with a higher level of achievement will have less difficulty meeting the proficiency standard than those who enter with a lower level. (Specifically, the former may have already met the standard or may be very close to it, so they need

to make little or no progress to contribute to the school's target.)

Moreover, AYP comparisons are confounded with differences between the cohorts in successive years — differences that may have nothing to do with the schools being evaluated. For example, this year's entering fourth-graders may be more poorly prepared than last year's fourth-graders, making it more challenging for the school to meet its AYP target.

An alternative view, while recognizing the importance of setting a single goal for all students, holds that meaningful and defensible judgments about teachers or schools should be informed by their contributions to the growth in student achievement and not based solely on the proportions of students who have reached a particular standard. In other words, only by following individual students over time can we really learn anything about the roles of schools and teachers.<sup>6</sup> That seems common-sensical — and VAM appears to make it feasible.

For that reason, many individuals and organizations have seized on VAMs as the “next new thing.” Many reports, as well as articles in the popular press, tout VAMs as the best, if not the only, way to carry out fair teacher evaluations.<sup>7</sup>

Such widespread interest in VAMs, not to mention their adoption in a number of districts and states, has spurred technical reviews.<sup>8</sup> The reviews paint a

---

5. R. L. Linn, “Assessments and Accountability,” *Educational Researcher*, vol. 29, no. 2 (2000), pp. 4-14. For a perspective on the experience in England, see L. Olson, “Value Lessons,” *Education Week*, vol. 23, pp. 36-40, May 5, 2004.

6. M. S. McCall, G. G. Kingsbury, and A. Olson, *Individual Growth and School Success*, Lake Oswego, OR: Northwest Evaluation Association, 2004; R. L. Linn, *Rethinking the No Child Left Behind Accountability System* (Paper presented at the Center for Education Policy, No Child Left Behind Forum, Washington, DC, 2004); H. C. Doran and L. T. Izumi, *Putting Education to the Test: A Value-Added Model for California*, San Francisco, CA: Pacific Research Institute, 2004; and D. R. Rogosa, “Myths and Methods: Myths About Longitudinal Research, Plus Supplemental Questions,” in J. M. Gottman (Ed.), *The Analysis of Change*, pp. 3-66, Hillsdale, NJ: Lawrence Erlbaum Associates, 1995.

7. D. Fallon, *Case Study of a Paradigm Shift (The Value of Focusing on Instruction)*, Education Commission of the States, Fall Steering Committee Meeting, Nov. 12, 2003; K. Carey, “The Real Value of Teachers: Using New Information About Teacher Effectiveness to Close the Achievement Gap,” *Thinking K-16*, vol. 8, pp. 3-42, Education Trust, Winter 2004; A. B. Bianchi, “A New Look at Accountability: ‘Value-Added’ Assessment,” *Forecast*, 1(1), June 2003; K. Raffaele, *Schools See ‘Value-Added’ Test Analysis as Beneficial*, Retrieved March 19, 2004, from the online edition of the *Pittsburgh Post-Gazette*, 2004; K. Haycock, “The Real Value of Teachers: If Good Teachers Matter, Why Don’t We Act Like It?” *Thinking K-16*, 8(1), pp. 1-2, *Education Trust*, Winter 2004; and D. M. Herszenhorn, “Test Scores to Be Used to Analyze Schools’ Roles,” *New York Times*, June 7, 2005, p. B3.

8. R. Bock, R. Wolfe, and T. Fisher, *A Review and Analysis of the Tennessee Value-Added Assessment System* (Technical Report), Nashville, TN: Tennessee Office of Education Accountability, 1996; R. Meyer, “Value-Added Indicators of School Performance: A Primer,” *Economics of Education Review*, vol. 16, pp. 183-301, 1997; H. Kupermintz, “Teacher Effects and Teacher Effectiveness: A Validity Investigation of the Tennessee Value Added Assessment System,” *Educational Evaluation and Policy Analysis*, vol. 25, pp. 287-298, 2003; and McCaffrey et al., 2003.

somewhat different picture. While acknowledging that VAMs are an important advance in applied measurement, the reviewers all strongly caution against their uncritical application, especially if the results are to have serious consequences for individuals or schools. Ultimately, the concerns are related to questions of proper test use.<sup>9</sup>

## 2. What Is the Fundamental Concern About VAMs?

The application of most VAMs involves both intricate statistical methodology and knotty questions of interpretation. But before confronting some of the technical issues arising in the use of VAMs in teacher evaluation, it is important to raise a fundamental problem that bedevils any attempt to measure teacher effectiveness.

At the conclusion of a value-added analysis, a number is associated with each teacher. That number, expressed in scale score points, may take on both positive and negative values. It describes how different that teacher's performance is from the performance of the typical teacher, with respect to the average growth realized by the students in their classes. It is often referred to as a measure of "teacher effectiveness." A problem arises because the word "effectiveness" denotes a causal interpretation. That is, the reader is invited to treat those numbers as if, in fact, they unambiguously capture the relative contributions of different teachers to student learning. Thus, if a teacher with an effectiveness of +6 were replaced by a teacher with an effectiveness of only +2, we should expect that the test scores in a typical class would be lower by an average of four points, other things being equal.

Obviously, such a change can never be directly observed because the same class cannot be simultaneously taught full time by two different teachers. So we must somehow infer, from the data we do have, what the relative contributions of different teachers

would be. To make the causal interpretation explicit, we have to specify the populations under study, describe the nature of the measure(s) employed, and define effectiveness in precise, quantitative terms.<sup>10</sup>

For example, we might want to evaluate all fourth-grade teachers in a particular district, using as our measure the increase in scores on a particular test over the course of the school year. We could define the effectiveness of a teacher as the difference between the average gain that would be obtained by a class taught by this teacher and the average gain that would be obtained by that same class if taught by the average teacher in the district. This would constitute a comparative or relative approach to teacher evaluation.

According to statistical theory, the ideal setting for obtaining proper estimates of teacher effectiveness (as defined above) is a school system in which, for each grade, students are randomly grouped into classes, and teachers in that grade are randomly allocated to those classes. Roughly speaking, randomization levels the playing field for all teachers in that each teacher has an equal chance of being assigned to any class.<sup>11</sup> The data generated in such a setting would allow us to obtain a reasonable estimate of each teacher's effectiveness, as well as a measure of the precision to be attached to the estimate. A finding that the average student growth associated with a particular teacher is significantly greater than the district average would be credible evidence for that teacher's relative effectiveness.

Unfortunately, school systems do not operate by randomization. Many parents have strong opinions about which districts (and even which schools within districts) they want their children to attend, and make corresponding decisions about housing. Within a school, parents often exert influence on the class or teacher to which their child is assigned. Similarly, teachers can sometimes select which district to work in and, by dint of seniority, have some choice in the

9. American Psychological Association, American Educational Research Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing*, Washington, DC: American Psychological Association, 1999.

10. These issues are explored in greater detail in McCaffrey et al., 2003, pp. 7-15.

11. Although randomization is an essential component of a proper experiment, there are additional complications in the teacher evaluation setting. See the answer to question 3 for more details.

classes they teach, or even the schools in which they are placed.

Since randomization is typically infeasible for the purpose of estimating teacher effects, we must resort to collecting data on teachers and students as they are found in their schools and classrooms. We then use statistical models and procedures to adjust, to the extent possible, for circumstances such as those just described.<sup>12</sup> It is impossible, however, to document and model all such irregular circumstances; yet they may well influence, directly or indirectly, the answers we seek nearly as much as what the teacher actually does in the classroom.

The fundamental concern is that, if making causal attributions is the goal, no statistical model, however complex, and no method of analysis, however sophisticated, can fully compensate for the lack of randomization.

The problem is that, in the absence of randomization, it is hard to discount alternative explanations for the results. (This explains why many consider randomized experiments the gold standard in scientific work.<sup>13</sup>) Specifically, teacher effects based on statistical estimates may actually represent the combined contributions of many factors in addition to the real teacher contribution we are after. Thus the estimate could be fundamentally off target.<sup>14</sup> Further, it is usually difficult to determine how off target an estimate is. Clearly, substantial discrepancies would seriously undermine the utility of inferences made on the basis of the analysis.

A number of authors have highlighted the distinction between “effects,” which are the output of a statistical algorithm, and “effectiveness,” which is an interpretation relating to the direct contribution of a teacher to student academic growth.<sup>15</sup> Careful consideration of that distinction in the context of schools brings to the fore the many pitfalls in

interpreting “effects” as “effectiveness.” That is exactly where the lack of randomization causes difficulties.<sup>16</sup>

Developers of VAM software and those who employ the results rarely acknowledge the implications of the fundamental problem. The assumptions required to justify endowing the estimated teacher effects with a causal interpretation (i.e., treating them as statistically unbiased estimates of teacher effectiveness) are usually not made explicit. Simply said, VAM proponents are behaving as if any statistical bias is too small to worry about. Unfortunately, most of the assumptions made are not directly testable. Thus, the credibility of the causal interpretations, as well as the inferences and actions that follow, must depend on the plausibility of those assumptions. In the context of real-world schools, judging plausibility is a very difficult matter.

### 3. What Are Some Specific Concerns About Treating Estimated ‘Teacher Effects’ as Measures of ‘Teacher Effectiveness’?

#### Inappropriate attribution

Because the ways teachers and students are matched in real schools may be related to the students’ potential or rate of growth, teachers can be inappropriately credited or penalized for their students’ results. For example, teachers with seniority are usually given more choice in the schools and classes they teach. Suppose they opt to work in schools with better conditions and in classes with students who are better prepared and more engaged. Those students may have a greater intrinsic rate of growth and, consequently, their teachers’ (apparent) effectiveness could be inflated. Conversely, newer and less-qualified teachers may be assigned to schools with poorer conditions and to classes with

12. In this context, the use of (simple) average gain scores cannot be recommended. More complicated methods are called for.

13. This point is somewhat controversial. A good general presentation can be found in R. J. Shavelson and L. Towner, *Scientific Research in Education*, Washington, DC: National Academy Press, 2002. For discussion, see M. J. Feuer, L. Towner, and Richard J. Shavelson, “Scientific Culture and Educational Research,” *Educational Researcher*, vol. 31, no. 8 (2002) pp. 4-14; and J. A. Maxwell, “Causal Explanation, Qualitative Research, and Scientific Inquiry in Education,” *Educational Researcher*, vol. 33, no. 2 (2004), pp. 3-11.

14. The technical term for an estimator being off target is that it is “statistically biased.” The use of the word “bias” here is different from such everyday meanings as “unfair” or “prejudiced.” Rather, it signifies that the differences between the estimator and its target cannot be made to vanish simply by accumulating more data.

15. For a lucid exposition in the present context, see Kupermintz, 2003.

16. Random matching of teachers and students would enable us to discount a number of alternative explanations for a finding of wide variation among estimated teacher effects. See Shadish et al., 2002.

students who are less prepared and less engaged. The (apparent) effectiveness of these teachers will likely be reduced. VAMs generally cannot eliminate these systematic misattributions.

Another, related, issue concerns so-called context effects. Student learning during the year is not just a function of a student's ability and effort, and the teacher's pedagogical skills. It also is affected by such factors as peer-to-peer interactions and overall classroom climate. To be sure, those variables are partially affected by the teacher; but with VAMs, the estimated teacher effect fully incorporates the contributions of all those factors, because no other component of the model captures them. That can also be a source of misattribution.

Further, student learning can be influenced by characteristics of the school, such as the availability of school resources, as well as by both school policies and differential treatment of schools by the district. Because teachers are not randomly distributed across schools, if those factors are not included in the model, then their contributions to student learning are absorbed into the estimated teacher effects.<sup>17</sup>

There is no easy way to address the issues. Including a school model in the VAM system can help somewhat, but may introduce other biases when there is a clustering of teachers of (true) differential effectiveness by school.<sup>18</sup> To fully disentangle the contributions of the different factors in order to isolate the teacher's contribution (i.e., obtain a statistically unbiased estimate of a teacher's effectiveness) is essentially impossible.

### Consequences of missing data

A district database compiled over time will generally have a substantial amount of missing data. Most commonly, the link between a student and a teacher for a given subject and grade is missing. If

there are student test data, they can be included in the calculation of the district averages but will not contribute directly to the estimation of teacher effects. If that link and/or the test score are missing is related to the score the student received, or would have received, then there is some bias in the estimated teacher effects.<sup>19</sup>

### Assumptions underlying the models

Another set of problems arises whenever one relies on mathematical models of real-world phenomena. In one VAM version, for example, it is assumed that a teacher's effect is essentially the same for all of that teacher's students in a given subject and year and, moreover, that the effect persists undiminished into the future for those students. Such assumptions may be more or less plausible, but they do require some validation rather than being accepted uncritically. If these assumptions substantially deviate from reality, the resulting estimates of teacher effects will be biased.<sup>20</sup>

In comparing teachers in a particular grade on the basis of their estimated effects, there is an implicit assumption that they all have been assigned similar academic goals for their classes and have equivalent resources. This flies in the face of the reality that tracking is endemic in schools, particularly in middle schools and above. Students in different classes may be exposed to different material and assigned different end-of-year targets. Those differences will influence the estimates of teacher effects. Moreover, different schools in the same district may be employing different curricula or following different reform strategies.

Another critical decision centers on whether to incorporate student background characteristics in the model. As we shall see, some approaches to VAM do and some don't. Although student characteristics are strongly correlated with student attainment, it

17. It even appears that students do more poorly in a grade if it is the lowest grade in the school. So, for example, seventh-grade students in a school with only seventh and eighth grades do more poorly on average than students in a school with sixth, seventh, and eighth grades. See W. J. Sanders and S. Horn, "Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research," *Journal of Personnel Evaluation in Education*, vol. 12, (1998), pp. 247-256. Again, these contextual factors affect the estimated teacher effects.

18. McCaffrey et al., 2003.

19. The technical term is that the missing data are not missing at random. See McCaffrey et al., 2003, for a discussion of the problem.

20. McCaffrey et al., 2003, develop a model that does not require the assumption that teacher effects persist undiminished into the future. When they apply this model to data, they obtain different results from those obtained with the assumption. However, there is an argument for assuming the simpler model, based on the relationship between the estimated effects and student characteristics. (W. J. Sanders, personal communication, March 18, 2005.)

appears that the correlation is much weaker with changes in attainment. That is an empirical finding, and it may still be the case that bias can be introduced if the model does not capture certain features of the students' demographics.

### Precision of estimates

The randomized experiment is considered to be the gold standard in experimental work. In medicine, for example, randomized clinical trials are almost always required to obtain FDA approval for a new drug or procedure. In such cases, however, the number of alternative treatments to be compared is usually rather small — typically fewer than four. If a large number of patients have been randomly allocated to the different treatments, we can assume that other factors besides the treatments under investigation that might affect the study outcomes have been averaged out across the treatments. Consequently, a statistically significant observed difference between a pair of treatments can be reasonably attributed to a real difference in efficacy, because plausible alternative explanations are unconvincing.

Unfortunately, obtaining useful estimates of teacher effects is more problematic — even if random allocation were feasible. The difficulty is that, in the education setting, teachers play the role of treatments. Thus, in a typical district with hundreds of teachers, the amount of information available for each teacher is relatively small, consisting of the data from just a few classes. Some VAMs try to remedy the situation by augmenting the data available for each student by including test scores from previous and future years, as well as from different subjects. While this can help, it does raise other concerns, as we shall see.

With a relatively small number of students contributing to the estimated effect for a particular teacher, the averaging power of randomization can't work for all teachers in a given year. Suppose, for example, a small number of truly disruptive students are in a cohort. While all teachers may have an equal chance of finding one (or more) of those students in their class each year, only a few actually will — with potentially deleterious impact on the academic

growth of the class in that year. The bottom line is that even if teachers and students come together in more or less random ways, estimated teacher effects can be quite variable from year to year.

In summary, given sufficient data, a reasonable statistical model, and enough computing power, it is always possible to produce estimates of what the model designates as teacher effects. Those estimates, however, capture the contributions of a number of factors, those due to teachers being only one of them. Moreover, the estimates may be quite volatile. So treating estimated teacher effects as accurate indicators of teacher effectiveness is problematic. Much more needs to be known about those kinds of data and the properties of the models in different, commonly occurring situations before there can be agreement on whether it is generally possible to isolate teachers' contributions to student learning, and have the confidence to carry out actions on that basis.

## 4. What Value-Added Models Are Now in Use?

Several VAMs are in circulation. They are similar in that they are purely statistical in nature and rely solely on student test scores, and not on other measures of student learning or such sources of information as interviews with students, teachers or administrators. Users of any of those models must confront the fundamental problem that the lack of random pairings among students and teachers makes causal attributions very problematic. The models do differ, however, in their structure and the kinds of assumptions they make.<sup>21</sup> The outcome of applying any of those models is that some number of teachers are identified as being significantly better or worse than average. Not surprisingly, findings can differ across approaches. Some VAMs are listed below:

- EVAAS (the Educational Value-Added Assessment System) is the best known and most widely used VAM. It was developed by William Sanders and his associates for use in Tennessee and has been in place there since 1993. Since then, it has been considered and, in some cases, adopted

---

21. McCaffrey et al., 2003.

by districts in other states.<sup>22</sup> (A fuller description of the EVAAS can be found in the answer to Question 5.)

- DVAAS (the Dallas Value-Added Accountability System) is a widely cited alternative to the EVAAS and has been employed by the Dallas school system for a number of years.<sup>23</sup> It uses a value-added criterion to identify highly effective teachers, as well as those in need of support. The DVAAS differs from the EVAAS in four important ways. First, it does use student-level characteristics to adjust student test scores prior to analysis. Second, it only models the relationship between adjusted test scores in adjacent grades (as opposed to combining data across several grades). Third, it doesn't directly model gains in the adjusted test scores but, rather, a more general structural connection between them.<sup>24</sup> Finally, the model includes not only a teacher's contribution to student achievement but also a number of other factors that are intended to account for the influence of the school on student achievement.
- Another alternative, REACH (Rate of Expected Academic Change), has been suggested by Doran and Izumi for use in California. Their test-based criterion measures student progress toward meeting a proficiency standard. Thus, each student's growth is measured against a goal rather than against the growth of other students. Doran and Izumi argue that this is a more constructive way of measuring AYP. They note that such a

value-added criterion could also be used to evaluate teacher effectiveness, but do not suggest a particular model for obtaining estimates of teacher effects.<sup>25</sup>

Other VAMs have been proposed, but used only in a research context.<sup>26</sup> Since EVAAS is the most widely used model for evaluating teacher effectiveness, this paper will henceforth focus on that model.

## 5. How Does EVAAS Work?

The building blocks of the EVAAS model are rather simple, with the complexity arising in the aggregation of data across students, subjects and years. We confine ourselves here to a necessarily incomplete summary.<sup>27</sup> The basic model is an equation that expresses the score of a student at the end of a particular grade in a particular year as the sum of three components:

- District average for that grade and year
- Class (teacher) effect
- Systematic and unsystematic variations

Thus, the essential difference between the student's score and the average score in the district is attributed to a "class effect" plus the combined contributions of unspecified variations, including measurement errors. It is assumed that the class effect is the same for all the students in the class and attributable to the teacher of the class.<sup>28</sup> (Henceforth, we refer to it as a teacher effect.) When the student moves to the next year and the next grade, the

22. The Ohio Partnership for Accountability, including all 51 of Ohio's schools of education, the State Department of Education, and the Board of Regents, has announced a project to use value-added teacher effectiveness data to better understand, study and improve university teacher preparation programs. Ohio is using a variant of the EVAAS. With cooperation of the state's teachers unions, Ohio's project is the first statewide effort of its kind. Over the next five years, Ohio researchers will study the math and reading scores of the students of both new and veteran teachers as a means to evaluate the quality of teacher preparation and to identify the most effective practices and policies. The Milken Foundation's Teacher Advancement Program, operating in schools in Arizona, Colorado, Indiana, Louisiana and South Carolina, includes the value added by teachers' achievement in its school reform model. Other districts using value-added measures of teacher effectiveness to improve teaching and learning include the Minneapolis Public Schools, Guilford County, North Carolina, as well as a number of districts in Pennsylvania.

23. W. Webster and R. Mendro, "The Dallas Value-Added Accountability System," in J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*, pp. 81-99. Thousand Oaks, CA: Corwin Press, Inc, 1997.

24. The technical term for this type of model is "analysis of covariance."

25. Doran and Izumi, 2004.

26. B. Rowan, R. Correnti, and R. J. Miller, "What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights From the Prospects Study of Elementary Schools," *Teachers College Record*, vol. 104, (2002), pp. 1525-1567.

27. The best technical description of EVAAS can be found in W. L. Sanders, A. Saxton, and B. Horn, "The Tennessee Value-Added Assessment System: A Quantitative Outcomes-Based Approach to Educational Assessment," in J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*, pp. 137-162, Thousand Oaks, CA: Corwin Press, Inc, 1997.

28. The identification of the class effect with teacher effectiveness conflates two separate steps: First, endowing a statistical quantity (class effect) with a causal interpretation and, second, attributing the causal contribution of the class entirely to the teacher. See H. I. Braun, "Value-Added Modeling: What Does Due Diligence Require?" in R. Lissitz (Ed.), *Value Added Models in Education: Theory and Applications*, pp. 19-39. Maple Grove, MN: JAM Press, 2005.

model then has four components:

- District average for that grade and year
- Teacher effect for that year
- Teacher effect from the previous year
- Systematic and unsystematic variations

Note that it is assumed that the teacher effect for the previous year persists undiminished into the current year and that the components of the unspecified variations in the two years are unrelated to each other. Finally, if we subtract the first-year score from the second-year score, we obtain the gain made by the student. According to the model, that must be the sum of:

- Average gain for that grade in the district
- Teacher effect of the second-year teacher
- The two error terms

That is, ignoring the error terms, the teacher effect in the second year is the difference between the gain experienced by the student in that year and the average gain in the district for that same year. This formulation is intuitively plausible and attractive.

It is possible to add equations for the data from subsequent years. Sanders uses the term “layered model” to capture the notion that the data from each succeeding year are added to those from the previous years. In a typical application, students may contribute as many as five years of data. Moreover, student gains in different subjects are included in the EVAAS model, with each subject and year assigned its own equation. It is not hard to see why the database to support the analysis is both large and complex, as it must maintain multiple links between students and teachers over different subjects and years.

The estimate of a teacher effect is based on many different elements, including the growth in learning (as measured by an increase in test scores) of the students in the teacher’s classes over a number of years, adjusted for the effects of previous teachers of those students; the growth of the teacher’s students

in subsequent years; and the achievements of those students in different subjects over a number of years, all appropriately adjusted for the contributions of those students’ other teachers.<sup>29</sup> It is virtually impossible to visualize how all these elements are combined to yield an estimate of the teacher’s value.

Sanders argues that there is no need to include student characteristics (e.g., gender, race, socio-economic status, and so on) in the model. His rationale is that, while there are substantial correlations between these characteristics and the current level of achievement, the correlations of these characteristics with gains are essentially zero. However, this is an assertion based on his reading of the data and not a mathematical certainty. This issue has been subjected to empirical examination and has not been found to be universally valid.<sup>30</sup> For this reason, some argue that fairer estimates of teacher effects will result if student characteristics are included in the model. Recently, it has been shown how this can be done as part of the EVAAS approach.<sup>31</sup> Unfortunately, this is not the end of the story, since the issues raised so far are still relevant to the proper interpretation of the resulting estimates.

The EVAAS model is very efficient in that it makes use of all the test information available for a given cohort of students within a moving five-year window. The estimation algorithms are able to handle various patterns of missing data so that if data on a particular student are unavailable in a given year, the remaining data can be incorporated into the analysis. In particular, student data that are not linked to a specific teacher still contribute to the estimation of the district averages. Sanders is correct in citing this as an advantage of his approach. However, as indicated earlier, if the patterns of missing data are related to student performance or teacher effectiveness, then systematic errors can be introduced into the estimated teacher effects. Sanders claims that by incorporating information over time and across subjects, the estimates generated by the

29. Although there is a separate equation for each subject and year, all the equations for a given cohort are tied together through another model feature (covariance matrices) that captures the fact that test scores for a given student over time and across subjects are statistically related to one another. This knitting together of disparate test scores distinguishes EVAAS from approaches based on simple comparisons of average gains across classes.

30. McCaffrey et al., 2003.

31. D. Ballou, W. Sanders, and P. Wright, “Controlling for Students’ Background in Value-Added Assessment for Teachers,” *Journal of Educational and Behavioral Statistics*, 2004.

EVAAS model are relatively unaffected by unusual patterns of missing data. Again, this claim requires empirical validation.

The principal output of an EVAAS analysis is a set of estimated teacher effects.<sup>32</sup> These estimates have well-established statistical properties. From the various studies they conducted, Sanders and his associates observed some heterogeneity among the estimated effects, which they interpreted as indicating real differences in teacher effectiveness.<sup>33</sup> (Indeed, they argue, as do many others, that teachers are the main source of variation in student gains.) Empirically, however, no more than a third of the teachers in a district have been reliably shown by EVAAS to be different from the average. Often the fraction is much smaller.

## 6. What Are Some of the Issues in Using Student Achievement Data in Teacher Evaluation?

It seems quite reasonable to judge teachers on the basis of their contributions to student learning.<sup>34</sup> Operationally, this means relying on scores obtained from standardized tests. One of the attractive properties of those scores is that they are hard numbers, as opposed to other qualities of students that we might be interested in documenting, such as engagement and enthusiasm, which are more difficult to measure.

We should recognize, however, that test scores are the final result of a complex process that involves translating state standards into test specifications and those, in turn, into test items assembled in a particular way to constitute an operational assessment. At each stage, design decisions are made on the basis of professional judgment, balancing substantive and psychometric considerations against constraints of

cost, testing time, and so on. Good practice requires that such test characteristics as the nature of the scale score and the validity of the test be examined in light of the proposed uses of the test scores.

**The Score Scale.** If different forms of a test are used for a particular grade each year, as is usually recommended, then the scores in the same grade from different years must be put on the same scale so that gains in different years are comparable. That involves a statistical procedure called (horizontal) equating that is common practice. Though it is usually done well, it does introduce uncertainty into the reported scores.<sup>35</sup>

Of greater concern is that, in some applications of VAMs, student scores over as many as five grades may be included in the database. Those scores are not obtained from a single test form administered in all the grades but from a number of test forms that, presumably, have each been designed to be grade-appropriate. Consequently, as we move to higher grades, the detailed specifications that govern the construction of each test will reflect the greater dimensionality and expanded knowledge base of the subject. This evolution in complexity is masked somewhat because test results are summarized in a single total score, and secondary analyses for evaluation typically utilize this total score.<sup>36</sup>

The total scores on the different instruments are usually placed on a common scale through another statistical procedure called (vertical) scaling, which introduces additional uncertainty into the process.<sup>37</sup> Aside from the technical aspects of vertical scaling, there is a question of what it means to put, say, third- and seventh-grade mathematics scores on the same scale. In particular, should we treat a 20-point gain at the low end of the scale as equivalent to a

32. For a given year and subject, a teacher can be associated with as many as three estimated effects, one for each of three successive cohorts. In Tennessee, schools are provided with the average of these three effects. In the following year, the data window shifts: the earliest cohort is dropped, effects for the two remaining cohorts are reestimated, an effect for a new cohort is obtained, and a new, three-cohort average is calculated.

33. W. L. Sanders and J. C. Rivers, *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*, Knoxville, TN: University of Tennessee Value-Added Research Center, 1996; and Sanders and Horn, 1998. Estimated teacher effects are normatively defined; i.e., as deviations from the average teacher in the district. As such, teacher effects cannot be compared across districts that were separately analyzed.

34. Educational Testing Service, *Where We Stand on Teacher Quality* (Teacher Quality Series), Princeton, NJ: Educational Testing Service, 2004.

35. M. J. Kolen and R. L. Brennan, *Test Equating: Methods and Practices*, New York, NY: Springer, 1995.

36. The total score is, typically, not a simple sum of the number of correct responses. It is, rather, a weighted composite constructed from subscale scores derived in turn from complex measurement models applied to the raw test data.

37. D. J. Harris et al., *Vertical Scales and the Measurement of Growth*, paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA, 2004. McCaffrey et al., 2003, also discuss some of the measurement issues. For a more accessible discussion, see D. Ballou, "Sizing Up Test Scores," *Education Next*, 2002. Retrieved May 26, 2004, from [www.educationnext.org/20022/10.html](http://www.educationnext.org/20022/10.html).

20-point gain at the upper end? Doing so requires making very strong assumptions about the nature of growth over the grade span of interest. More to the point, should we expect that the average teacher teaching a typical class would obtain the same (relative) growth irrespective of the grade? That is unlikely given the increasing complexity of the construct at higher grades. Although the question could be addressed empirically, that appears not to have been done.

**Validity.** Whether test scores actually measure what they are intended to measure is the basic concern of validity.<sup>38</sup> Typically, state content standards are broad, ambitious and often ambiguous. The degree of articulation between tests and the standards varies among states and even across subjects and grades within a state. Indeed, reviews of state tests often find that they don't measure some of the content standards at all and some only superficially, focusing instead on those aspects of the standards that can be probed with multiple-choice questions.<sup>39</sup> For example, a standard in language arts addressing the ability of a student to write a well-crafted essay should be measured by having the student write an essay. Most would agree that a multiple-choice test falls short. Those considerations give rise to two related concerns: First, that "teaching to the test" may result in increased test scores that do not generalize to gains in the broader achievement domain that the test is intended to measure. Second, that teachers who do try to teach the full curriculum may find their students not gaining as much as others, whose teachers resort to some form of teaching to the test.<sup>40</sup> That is, the test may not be sensitive to the full range of students' learning gains. That problem possibly

is exacerbated by the use of test scores obtained through a vertical scaling procedure.<sup>41</sup>

In sum, a rigorous evaluation of the validity of the assessment battery used by a state is an essential foundation for appropriate test use. The alignment of the test with the corresponding standards, as well as the shift in the meaning of the score scale across grades, should be taken into account in deciding how to best use test scores. Given the current state of the art, caution is warranted. Policymakers should have technical support in deciding whether the test score scale can support the interpretive burden placed on it — and moderate their use of VAM results accordingly.

To cite one example, suppose that, on average, reported score gains are typically smaller the higher the students' initial scores. In that case, two teachers of equal effectiveness, but assigned over time to classes with substantially different distributions of initial scores, can find themselves with quite different estimated effects. If those are interpreted as indicators of differential effectiveness, then teachers are ill-served by the process. Of course, those difficulties would be mitigated somewhat if we only compared teachers at the same grade level.<sup>42</sup>

## 7. Where Do We Stand?

Progress and promise are found in that:

- VAM moves the discussion about teacher quality to where it belongs: centered on increasing student learning as the primary goal of teaching. It can also enhance the teacher evaluation process by introducing a quantitative component, as well as by forcing us to re-examine questions of fairness and proper test use. These are major steps in the right direction.

38. See for example L. J. Cronbach, "Five Perspectives on Validity Argument," In H. Wainer and H. I. Braun (Eds.), *Test Validity*, pp. 3-17. Hillsdale NJ: Lawrence Erlbaum Associates, Inc., 1988. The classic reference is S. Messick, "Validity," In R. L. Linn (Ed.), *Educational Measurement*, 3rd Ed., pp. 13-103. New York, NY: Macmillan Publishing Co., 1989.

39. American Federation of Teachers, "Executive Summary," *Making Standards Matter* 2001. Retrieved May 25, 2004, from [www.aft.org/edissues/standards/MSM2001/downloads/execsummary.pdf](http://www.aft.org/edissues/standards/MSM2001/downloads/execsummary.pdf). For an interesting perspective, see G. W. Bracey, *A Review of: The State of State Standards*, (Thomas B. Fordham Foundation, January 2000), (No. CERAI-00-07), Milwaukee, WI: Center for Education Research, Analysis, and Innovation, Feb. 2, 2000.

40. For two of many views, see L. Bond, "Teaching to the Test," *Carnegie Perspectives*, 2004. Retrieved Aug. 3, 2004, from [www.carnegiefoundation.org/perspectives/perspectives2004.Apr.htm](http://www.carnegiefoundation.org/perspectives/perspectives2004.Apr.htm), and W. J. Popham, "Teaching to the Test?" *Educational Leadership*, vol. 58, no. 6, (2001), pp. 16-20.

41. W. H. Schmidt, R. Houang, and C. C. McKnight, "Value-Added Research: Right Idea but Wrong Solution?," In R. Lissitz (Ed.), *Value-Added Models in Education: Theory and Applications*, pp. 145-164. Maple Grove, MN: JAM Press, 2005.

42. Some may think (wrongly, as it happens) that such difficulties are particularly severe with external tests. Actually, local tests can be deficient with respect to both reliability and validity. They are even more problematic with respect to comparability across schools or districts. Thus, there is probably no alternative but to use externally developed tests for teacher evaluation.

- By utilizing measures related to individual student growth, VAM provides a more defensible foundation for teacher evaluation than is offered by methods based on the proportion of students meeting a fixed standard of performance.
- Investigations of different VAMs in a variety of settings have begun to give us a clearer picture of the strengths and limitations of the various approaches.

Appropriate uses of VAM results exist, such as:

- Identifying teachers who are most likely to require professional development and who should be interviewed and/or observed to determine the particular kinds of support that would be most helpful. That screening strategy would help allocate scarce resources to teachers in greatest need.<sup>43</sup>
- Identifying schools that may be underperforming and should be audited to determine whether they are in need of specific kinds of assistance.<sup>44</sup>

Cautions remain, such as:

- VAM results should not serve as the sole or principal basis for making consequential decisions about teachers. Many pitfalls exist in making causal attributions of teacher effectiveness on the basis of the kinds of data available from typical school districts. We still lack sufficient understanding of how seriously the different technical problems threaten the validity of such interpretations.

- Although we can all agree on the importance of teacher evaluation, identifying precisely which teachers are deserving of commendation and which are in need of focused professional support is another matter entirely. Unfortunately, extreme ranks, those near the top or near the bottom, are very unreliable.<sup>45</sup>
- The use of VAMs should not block the examination of the appropriateness or desirability of including other measures, in addition to student test scores, in teacher evaluation.<sup>46</sup> Moreover, we must recognize that statistical models cannot identify the strategies and practices teachers employ. Expert observation, portfolio reviews, conversations with teachers, and so forth, are essential to making informed judgments about whether one teacher truly excels or whether another really needs support. School leaders should also become more skilled in recognizing the kinds of assistance needed by individual teachers.

43. This type of use was cited in Sanders and Horn, 1998, as the primary function of EVAAS and has been carried out in some districts in Tennessee. Apparently, Dallas has also made good use of its VAM results in building system capacity through targeted professional development. In this regard, see Webster and Mendro, 1997.

44. Underperforming schools could be identified by looking for clusters of teachers with low estimated effectiveness or by carrying out school-level value-added analysis.

45. J. Lockwood, T. Louis, and D. F. McCaffrey, "Uncertainty in Rank Estimation: Implications for Value-Added Modeling Accountability Systems," *Journal of Educational and Behavioral Statistics*, vol. 27, (2002), pp. 255-270.

46. In general, the use of multiple sources of information is preferable to the use of a single measure. At the same time, it is important to note that test scores have been subject to much greater scrutiny, and their properties more thoroughly documented, than have other possible measures such as those obtained through direct observation.

As a nation, we have come to the realization that good teaching really does matter. Building a credible statistical basis for teacher evaluation would be an important advance that could contribute, in the long run, to improved teaching and learning. But the evaluation process should be sufficiently rich to do justice to the complex, multifaceted activity that constitutes teaching. Reliance on a single, statistical measure cannot be recommended. That said, VAMs certainly play an important role. Indeed, the logic behind using VAMs is compelling, and VAM-based approaches to teacher accountability have gained numerous adherents in many states. In view of the methodological issues that have surfaced, however, further investigations of various approaches are critical. Fortunately, there is substantial activity in this area.<sup>47</sup>

It may well be that we can never rigorously justify treating estimated teacher effects as accurate indicators of teacher contributions to student learning. Nonetheless, districts employing VAM results in sensible ways might, over time, experience greater improvements in student scores than other comparable districts not using VAMs. That possibility can be investigated using a randomized experiment conducted at the district level. Indeed, it has been argued that such a study would yield results that are more directly related to policymaker concerns than are attempts to validate the causal interpretation of VAM output.<sup>48</sup>

Policymakers should not ignore the technical aspects of VAMs. The concerns that have been raised are central to the proposed use of VAM results in teacher evaluation. An early objection to the EVAAS system was that it was too difficult to understand and thus shouldn't be used to make decisions about

teachers. The response was that one didn't have to understand how a car works in order to drive it. That argument seemed to carry the day. However, in view of more recent critiques, only summarized here, perhaps the metaphor should be re-examined.

Certainly, one needn't understand how a car works while driving it under the conditions it was designed for. But if one plans to drive it under nonstandard conditions, say on a beach, it is only prudent to inquire first about the capabilities of the drive train and the transmission before setting off. Similarly, the statistical models underlying EVAAS were originally developed for use in settings, such as agriculture, in which randomized experiments and sufficient data are the norm. Thus, endowing statistical estimates with causal interpretations is relatively straightforward. But taking that same methodology off-road, so to speak, in circumstances with multiple sources of selection bias (and perhaps less data than desired), demands a careful look under the hood. That is just due diligence.

Finally, raising the quality of teaching will require more than instituting better accountability. At the least, jurisdictions implementing VAMs should also be building capacity to help those teachers who are identified as needing improvement. But greater effort is called for: States and districts have many levers at their disposal, among them the improvement of teacher training, standards for licensure, effective mentoring for new teachers, policies regarding the assignment of teachers to schools and to classrooms within schools, more equitable distribution of resources, targeted professional development, as well as higher salaries and differentiated pay schedules. A coherent, sustained and systemic initiative that involves many, if not all, of those levers will surely meet with greater success than a narrow effort focused on just one.

---

47. See, for example, H. Wainer, "Introduction to the Value-Added Assessment," *Journal of Educational and Behavioral Statistics* (Special Issue), Vol. 29, (2004), pp. 1-3; and R. Lissitz (Ed.), *Value-Added Models in Education: Theory and Applications*, Maple Grove, MN: JAM Press, 2005.

48. See D. B. Rubin, E. A. Stuart, and E. L. Zanutto, "A Potential Outcomes View of Value-Added Assessment in Education," *Journal of Educational and Behavioral Statistics*, Vol. 29, (2004), pp. 103-116.

# ETS ON THE ISSUES

## It's Time to Act on Teacher Quality

By Kurt Landgraf, President and CEO, ETS

---

*Our nation is at a crossroads. We will not continue to lead if we persist in viewing teaching — the profession that makes all other professions possible — as a second-rate occupation.*

That warning comes from the recent report of the Teaching Commission. At ETS, we take the warning seriously.

Everything we are doing in the area of teacher quality starts with the belief that good teachers produce good students. Good teaching requires that both new and experienced teachers possess:

- thorough content knowledge
- knowledge of pedagogy
- demonstrated teaching skills

The formula for creating a highly qualified teaching force is neither secret nor complicated. It starts with rigorous standards for teacher preparation and licensure, and it continues with effective, sustained professional development throughout a career. And, to be as successful as we must be, we also need to look seriously at alternative school-staffing patterns and differentiated pay.

Unfortunately, the requirements for entry into the teaching profession are often too low. For example, only half of the country's teacher education institutions meet national accreditation standards. They all should. An ETS study, *Preparing Teachers Around the World*, found that countries with high-performing schools also have high, strictly regulated standards for licensure. That is not coincidental. No doubt, higher entry standards to the profession will also drive improvements in teacher preparation.

To ensure that teaching methods align with standards and curricula, teachers need more effective, ongoing professional development. School administrators play a critical role. They must sharpen their own skills as well.

On compensation and retention, the Teaching Commission argued: "Money matters." Public opinion research conducted for ETS confirms that Americans support higher teacher salaries even if it means paying higher taxes. Better pay will attract more highly skilled candidates to the profession, particularly to underserved and understaffed locations, subjects and specialties.

We must summon the political will to do whatever is necessary to ensure that every child in America is taught by a well-trained, well-qualified and well-compensated teacher.

ETS is committed to doing its part. Our partnership with the National Council for Accreditation of Teacher Education is setting national benchmarks for licensure tests. Our Recognition of Excellence award provides an incentive for candidates to excel on their licensure exams. And our *Praxis* Study Guides help candidates do their best on the tests, while our Diagnostic Preparation Program helps them improve their scores.

Where teacher quality is concerned, we're listening to educators, parents and policymakers. We're learning from sound research. And we're leading the effort to achieve both informed public policy and informed educational practice.

# Access and Equity: Understanding the Achievement Gap

## The Achievement Gap Develops Both in and Outside of the Classroom

An ETS report, *Parsing the Achievement Gap: Baselines for Tracking Progress*, identifies the factors before, during, and after school that create and perpetuate the well-documented gaps in achievement among students from different racial and ethnic backgrounds and different family income levels. Efforts to close the achievement gap will fail unless policymakers, educators, and parents recognize and address the variety of factors causing the gap, both in school and in society.

The ETS report specifies 14 factors related to student achievement across three main categories. The factors range from birth weight and hunger to lead poisoning, parental involvement and teacher quality, and are classified under the categories of early development, the school environment, and the home learning environment.

In every instance, the ETS report found minority students disadvantaged relative to White students in the conditions and experiences conducive to student achievement.

### The Achievement Gap Separates Communities by Race and Income

The achievement gap between racial and ethnic groups has widened. Black and Hispanic high school graduates are now performing in math and reading at the same levels as eighth-grade White students. Paul Barton, author of *Parsing the Achievement Gap*, writes that “The results are unambiguous. In all 14 correlates of achievement, there were gaps between the minority and majority student populations. Eleven of those also showed clear gaps between students from low-income families and higher-income families. The gaps in student achievement mirror inequalities in those aspects of school, early life, and home circumstances that research has linked to achievement.”

Early Development	The School Environment	Home Learning Environment
<ul style="list-style-type: none"> <li>• weight at birth</li> <li>• lead poisoning</li> <li>• hunger and nutrition</li> </ul>	<ul style="list-style-type: none"> <li>• rigor of the school curriculum</li> <li>• teacher preparation</li> <li>• teacher experience and attendance</li> <li>• class size</li> <li>• availability of appropriate classroom technology</li> <li>• school safety</li> </ul>	<ul style="list-style-type: none"> <li>• reading to young children</li> <li>• TV watching</li> <li>• parent availability and support</li> <li>• student mobility</li> <li>• parent participation</li> </ul>

## Everyone Is Accountable for Better Instruction and Improved Learning

Education reform requires both funding and accountability. Educators should be accountable for all their students, but the rest of us — elected officials, policymakers, parents and the public — must be held accountable to the pledge for better instruction and increased learning. We need to develop **strategies** that will provide a quality education to all students. We need to administer and score **educational assessments** appropriate to our students. We need to use **the results** of that testing to understand how well students are learning and to help educators to focus their time and resources on the keys to improvement. The challenges we face in education today will not end tomorrow, because they did not begin yesterday.

## ACCESS & EQUITY

---

*Historically, we have placed much greater emphasis on the need to provide equality of opportunity than on equality of outcomes. But equal opportunity requires equal access to knowledge. We cannot expect everyone to be equally skilled. But we need to pursue equal access to knowledge to ensure that our economic system works at maximum efficiency and is perceived as just in its distribution of rewards.*

— *Federal Reserve Chairman Alan Greenspan, in testimony before the House Education and Workforce Committee, March 11, 2004*

*The academic achievement gap is the major driver of racial inequity in this country. It is the unfinished business of the civil rights struggle.*

— *U.S. Secretary of Education Rod Paige*

*When assessing new reforms, it is important to separate fact from fiction. In fact, there is money in the bill to help teachers and assist schools in need of improvement, but many schools do not know what is available to them. For instance, under the President's 2004 budget, Colorado would receive \$104 million to raise student achievement in our high-poverty school districts.*

— *U.S. Sen. Ben Nighthorse Campbell, R-Colo., Senate Rural Education Caucus*

*I mean, in order to provide those types of things, whether it be people, resources or technology to bring kids up to speed or just dollars for other viable programs to put them at a level playing field, in my situation, it all starts with dollars and cents. And I don't care how you coin it, it's all the same.*

— *Woman, Columbus, Ohio, focus group for "Equity and Adequacy: Americans Speak on Public School Funding," a Hart-Teeter public opinion poll conducted in 2004 for ETS*

*Until America stops tying school funding so closely to local wealth, fair and equitable funding for real school reform will only be a dream.*

— *Kurt Landgraf, President and CEO, ETS*

*Why would you vote yes [on a school-budget increase] when every other time the tax has been increased it still doesn't go to education?*

— *Man without school-age children, Tampa, Fla., focus group for "Equity and Adequacy: Americans Speak on Public School Funding," a Hart-Teeter public opinion poll conducted in 2004 for ETS*

*Until we become more concerned about our neighboring schools, about all the students in the nation, until we get that global concern, instead of just what's best for me and my own little group right now, we are not going to have equity or adequacy for our children.*

— *Man, Columbus, Ohio, focus group for "Equity and Adequacy: Americans Speak on Public School Funding," a Hart-Teeter public opinion poll conducted in 2004 for ETS*

*I wouldn't mind giving an extra penny to sales tax, but I want to know where my money is going. You know, if you could prove that to me, the accountability again, it comes down to accountability. If they can prove it to me, then fine. I wouldn't mind doing it.*

— *Woman without school-age children, Tampa, Fla., focus group for "Equity and Adequacy: Americans Speak on Public School Funding," a Hart-Teeter public opinion poll conducted in 2004 for ETS*

# ONE-THIRD OF A NATION

*A Policy Information Perspective by Paul Barton*

---

## Executive Summary

---

This report is about high — and rising — high school dropout rates, some exemplary efforts to retain students, the limited — and diminishing — opportunities for dropouts to regain a footing in education and training, and the increasingly dire prospects for dropouts in today's economy. About a third of students are leaving high school without a diploma: one-third of a nation.

### The High School Completion Rate Has Not Been Accurately Reported

- Official estimates of state completion rates are too high, and the U.S. Department of Education is examining ways to obtain better measurements. One National Center for Education Statistics series going back to the 1880s is on the mark, but never seems to be reported in the press.
- A number of independent researchers have made recent estimates that put the national rate variously at 66.1, 66.6, 68.7, 69.6, and 71.0 percent.

### The High School Completion Rate Has Been Falling

- Nationally, after peaking at 77.1 percent in 1969, the rate dropped to 69.9 percent in 2000.
- From 1990 to 2000, the completion rate declined in all but seven states. In 10 states, it declined by 8 percentage points or more.

### The Completion Rates Vary Widely Among the States, in Close Relationship to Factors Identified by Research as Predictive of Students Dropping Out

- Recent completion rates range from a high of 88 percent in Vermont to a low of 48 percent in the District of Columbia and 55 percent in Arizona.

- A combination of three factors — socioeconomic characteristics, number of parents living in the home, and a history of changing schools frequently — are associated with 58 percent of the variation in completion rates among the states. Those combined factors predict estimated state completion rates within four percentage points in 24 states. A close study of the states with completion rates that are substantially different from predicted rates might reveal important differences that affect student retention. Such correlations do not predetermine school completion rates; efforts of various kinds to improve retention have shown results and are discussed in this report. Also, improvements in school quality that raise student achievement will also improve completion rates, for succeeding students are more likely to complete school.

### Ways to Increase Retention Have Been Demonstrated, Providing a Resource for School Systems to Follow

- Alternative schools, approximately 11,000 in number, have been established to serve students at risk of dropping out. But we need to know more about the schools that are successful and the different kinds of approaches they use.
- The Talent Development (TD) High School is a model reform program developed by the Center for Research on the Education of Students Placed at Risk. The TD high school is based on research on student motivation and teacher commitment. There are 33 TD high schools in 12 states, and evaluations have shown TD to increase student retention.

- Communities in Schools, a program run from a national office and five regional offices, has been around for a long time. Evaluations have shown that this type of program can increase student retention.
- Maryland's Tomorrow, begun in 1985, is a statewide dropout prevention program operating in 75 schools. Evaluations carried out by the state of Maryland and Johns Hopkins University have shown positive effects on student retention.
- The Quantum Opportunities Program was a program funded by the Ford Foundation with the intention of keeping students in school. The program had considerable success, running from 1989 to 1992, when the grant ended. The knowledge and experience developed by the program are still available.
- While those programs have helped us better understand some of the mechanisms that can help keep students in school, the homework on retaining students has not been finished; a need still exists for rigorous evaluation, and for replication of evaluations already conducted.

### **A Scarcity of Guidance and Counseling Personnel, and Related Staff, to Work One-on-One With Students at Risk of Dropping Out and Their Families, Exists**

- On average, only one certified counselor is available for each 500 students in all schools, and one counselor to 285 students in high schools.
- The ratio varies considerably. The ratio is higher in schools where less than half of the students are college bound; it is also higher in schools with a high percentage of minority students.
- The bulk of counselor time goes to helping students with the transition to college, getting students scheduled for classes, dealing with

student behavioral problems, and, increasingly, administering standardized testing as part of accountability systems. Little time is spent on career guidance for students going directly to work, or on transition-to-work services for such students. And little time is available for one-on-one work with students who are at risk of dropping out and their families. The counseling function has been largely ignored in the education reform movement.

### **Opportunities for Dropouts to Resume Education and Training Are Diminishing**

- The federal investment in second-chance programs has dropped from \$15 billion in the late 1970s, at a time when school completion was peaking, to \$3 billion today. Projects originating at the local level come and go, but national records are not kept on them.
- This decrease in program funding is happening at the same time that the earning power of dropouts is in sharp decline.

### **The GED Program Is an Important Second-Chance Opportunity, but Opportunities for Instruction in Second-Chance Programs Are Not Growing**

- While generally known as a program for adults, GED certificates have shifted toward 16- and 17-year-olds over the past decade. Several possible reasons are discussed in this report, but knowledge is incomplete.
- The American Council on Education has recently strengthened the GED and made it a more rigorous test. It is as yet unknown whether the new test will result in a change in the number of GEDs awarded.

## While Second-Chance Opportunities Have Diminished, Scientific Evaluations Have Shown Some Programs in Operation to Be Effective; a Base of Knowledge Exists for Rebuilding

- The Job Corps, with more than 60,000 enrollees, has been operating since 1964 and has been subject to many evaluations. The program has shown staying power and is the only such program operating that long.
- Youth Build USA provides education and training in the context of building affordable housing. Funded primarily by the Department of Housing and Urban Development, the program has created more than 12,000 housing units while training more than 40,000 youth.
- The Center for Employment Training (CET), begun in 1967, has 33 centers in 12 states. The CET provides job training and education. Evaluations have found the CET to be very effective.
- Youth Corps (Service and Conservation Corps) traces back to the Civilian Conservation Corps of the 1930s. After a period of federal funding ended, 31 states continued to operate programs, enrolling 23,000 youth annually. The program generates \$1.60 in immediate benefits for every \$1 of costs.
- The community college is a ubiquitous and flexible institution with a lot of involvement in GED and remedial instruction for dropouts. These colleges have the capability to make a much larger contribution. But unless special tuition reimbursement or grant programs are available, dropouts must pay tuition — often supporting themselves at the same time they are attending school.

## The Earning Power of High School Dropouts Has Been in Almost Continuous Decline Over the Past Three Decades

- High percentages of young dropouts are either not employed or are not even in the labor force. Most wander through life like lost travelers, without guidance or goals, and many end up in prisons.
- The earning power in constant 2002 dollars of 25- to 34-year-old dropouts who work full time for a full year has been in steady decline, during an age period critical to getting established, forming families, and raising children.
- In 1971, male dropouts earned \$35,087 (in 2002 dollars), falling to \$23,903 in 2002, a decline of 35 percent.
- In the same period, the earnings of female dropouts fell from \$19,888 to \$17,114.
- Earnings of high school graduates also dropped considerably, but not as much as earnings for those who drop out of school.

## The Nation Faces Increasing Dropout Rates, Declining Assisted Second-Chance Opportunities for Education and Training, and a Deteriorating Economic Position

- In high school completion rates, the United States has now slipped to 10th place in the world.
- Only the kind of national resolve being shown to raise student academic achievement can reverse these adverse trends for that third of the nation's youth. Increasing student achievement in the early years may well lead to increases in school completion since it is the low achievers who are more prone to dropping out.

To view the full report visit

[www.ets.org/Media/Research/pdf/PICONETHIRD.pdf](http://www.ets.org/Media/Research/pdf/PICONETHIRD.pdf)

# ETS ON THE ISSUES

## On Education Reform, Let's Avoid False Choices

*By Kurt Landgraf, President and CEO, ETS*

---

Policymakers face some of the most momentous issues in our nation's history. Education reform continues to be one of them.

From expanding No Child Left Behind to high school, to improving teacher quality and closing the achievement gap, public education is at a crossroads. The steps we take now will determine where we'll be in a generation.

Moving forward with NCLB will be a priority for the Secretary of Education. NCLB has highlighted soft spots in the system, thanks in part to data-collection provisions regarding student performance. That is precisely what NCLB was designed to do. Three years of experience have also shown us how NCLB might be implemented more effectively.

As debates over those issues take shape, we must avoid the traditional either/or formulation all too common in political and policy discussions: Either you're for holding schools accountable for student performance, or you're for more spending. Either you're for expanding opportunity to the many, or concentrating on the potential of the few. Either teachers should master content, or they should have better teaching skills.

Those are false choices that ill-serve students, parents and the public. At ETS, we think it is possible to pursue both accountability and efficiency; to pursue both equity and quality; to master both content and pedagogy.

Public opinion research conducted for ETS shows that Americans want reform to proceed. By more than two-to-one, parents say they think NCLB will improve public education. This insistence on reform has become the new middle ground in the education debate, and it forms the common ground on which policy should coalesce.

In the coming months, we'll be setting forth ETS's positions in three fundamental areas:

- ensuring that our schools are held accountable for student performance
- expanding access to educational opportunity
- improving teacher quality

By keeping the focus on education reform, and by avoiding the trap of false choices, we'll all be taking the right steps today for the generation of tomorrow.

At ETS, we're listening to educators, parents and policymakers. We're learning from sound research. And we're leading the effort to achieve both informed public policy and informed educational practice.

# What Will We Do When the Bell Sounds?

*By Kurt Landgraf, President and CEO, ETS*

---

When the bell rings in high schools around the country this morning, will our schools be equipped to deliver excellence? Recent trends are not encouraging.

Between 1990 and 2000, the high school completion rate in the United States fell 3.3 percent. We now rank 17th among developed nations in the percentage of youth graduating from high school.

Among the states, a yawning disparity on high school completion rates exists, from a high of 88 percent to a low of 48 percent.

Most troubling of all, the grades at which students are dropping out have shifted from 11 and 12 a generation ago, to grades 9 and 10 today. Dropouts are younger, less educated and less prepared for work, life and democracy than ever before.

That's the bad news. Here's the good news: Just as the No Child Left Behind Act is bringing tougher standards and accountability to grades K-8, there's a growing realization that we need to do the same for our secondary schools.

President Bush has made high school reform a centerpiece of his domestic agenda. In the states, Governor Mark Warner of Virginia, chair of the National Governors Association, is leading a drive to redesign the American high school.

"It is time for a new approach to high school," Governor Warner says, "one that both challenges our students more and gives them new opportunities to engage in meaningful, lifelong learning."

ETS is committed to high school reform. We think high schools need rigorous curricula, taught by highly qualified teachers overseen by school leaders able to attract and retain highly qualified faculty.

ETS has published research that examines high school completion rates. Our research serves as a warning on the status quo, a benchmark for progress, and a platform for a vigorous national debate on high school reform.

When the bell rings in high schools this morning, we must be prepared to deliver excellence.

At ETS, we're listening to educators, parents and policymakers. We're learning from sound research. And we're leading the effort to achieve both informed public policy and informed educational practice.

# The Public Speaks: On Education and Competitiveness

Results from the ETS 2006 Survey, *Keeping our Edge: Americans Speak on Education Reform and Competitiveness*, show that the American people recognize the important role that public schools play in contributing to our nation's achievements. They are concerned about our ability to remain globally competitive unless our schools improve and challenge students more. Despite major reform efforts at the national and state levels, public ratings for the nation's schools have not changed over the past six years of this survey series. The public feels strongly that America's schools must do a better job when it comes to preventing dropouts and raising standards, particularly in math and science, to ensure that high school graduates are prepared for college and for the increasingly technical jobs in the modern economy. A majority of the public expects that if changes are not made in our schools, negative effects on the economy will be felt within the next 10 years.

## **Math and Science are Crucial**

---

Math and science are viewed as vital to America's ability to maintain an edge in the global economy and remain competitive with nations such as China, India and Japan.

- The general public (40%) and opinion leaders (61%) identify math, science and technology skills as the most important ingredients in America's ability to compete in the global economy.
- More than two-thirds (72%) of Americans – and majorities of high school teachers and administrators, college faculty, and opinion leaders — believe that it is very important for students to take the most advanced math and science classes they are eligible to take, every year of high school.
- After hearing arguments stressing the importance of math and science, the proportion of the public believing it is important for students to take advanced classes increases to 76%. The strongest arguments stress the need to qualify for technical jobs in the modern economy and the need to respond to international competition.

## **A Call to Action**

---

While the public views math and science as vital to America's ability to maintain its edge in the global economy, Americans do not believe that our public

schools are providing students with the skills they will need to compete for highly technical scientific and engineering jobs.

- A 71% majority of Americans believe that our nation's public high schools are coming up short or falling behind in efforts to put students on the path to compete for highly technical scientific and engineering jobs with their counterparts from other countries.
- A 58% majority of Americans feel our nation's public schools are coming up short or falling behind in efforts to give students who want to go into the work force the training and skills they need to secure and succeed in a job.
- More than three-quarters (76%) of Americans believe that if America's next generation does not work to improve its skills in math, science, and engineering, it risks becoming the first generation of Americans who are worse off economically than their parents.
- Nearly two-thirds (64%) of Americans believe that if we fail to take action and reform our education system, our ability to remain globally competitive will be compromised within the next decade.
- And 73% of the public believes that if we fail to take action during the next 25 years, our complacency will negatively impact America's ability to compete worldwide.

## “Good Enough” Is Not Good Enough Accountability

---

Americans rate their communities’ public schools as somewhat better than average, but they are not complacent about this. All groups in the survey call for higher academic standards and endorse a wide range of potential reforms.

- When asked to assess America’s public schools as a whole, the public rates them as average (a grade-point average of 2.1, for a C grade). This number has barely budged over the six years of this survey.
- The assessment improves when parents, students, teachers, and administrators are asked to grade the schools in their community or their own school (grade-point averages ranging from 2.8 to 3.0, for a B grade), but this number, too, has hardly changed.
- Americans speak of reform in the future tense, saying that they believe public school reforms are absolutely vital to our ability to sustain a healthy, globally competitive economy. More than half (53%) of the public believes that major changes to or a complete overhaul of our public schools are needed — a proportion that almost matches the level in 2002.
- Only 11% of Americans feel that academic expectations are set high and that high school students are being significantly challenged in school.

The American public clearly sees room for improvement in the amount of effort that all parties — students, teachers, administrators, parents and the community at large — exert toward ensuring that students get the most out of their school experience. The public does not feel that any one particular group is putting a substantial amount of effort into students’ school experiences; yet they often characterize public schools as doing “well enough” along several dimensions.

- The public rates teachers and students as putting in more effort than any other groups, yet they only rate teachers and students as putting forth

a little more effort than “just enough to get by.” Only 14% of Americans feel that teachers are putting in as much effort as they can, and only 5% feel that students are putting forth maximum effort.

- Even when groups rate themselves (e.g., teachers rating teachers’ efforts), they assess their group’s effort as higher than other groups, yet no group rates themselves as exerting as much effort as they can.
- Americans are pretty evenly split on whether the nation’s public high schools are challenging and pushing the best students to make the most of their abilities — half believe that schools are doing at least well enough, and 46% feel that they are coming up short or falling behind in this area.
- College faculty (69% coming up short/falling behind) and opinion leaders (52%) are particularly concerned about how public high schools serve their brightest students.
- Americans also are split on whether the nation’s public high schools are preparing students to succeed in two- and four-year colleges. Forty-five percent believe that they are doing at least well enough, and 49% believe they are coming up short or falling behind.
- Forty-three percent of Americans feel that the nation’s public high schools are doing at least well enough at teaching the basics of math, science, and writing, while 55% believe that public high schools are coming up short or falling behind.
- Americans are of one mind when it comes to their assessment of how successful public schools are at supporting struggling students and preventing dropouts. Nearly three-quarters (73%) believe that they are coming up short or falling behind in these efforts.

## The Need for Inspirational Leadership

---

The public is eager for someone to take the reins, to step up and assume a leadership position, and ask all Americans — students, parents, teachers,

administrators, employers, the entire community — to help transform our public schools. Americans are not in denial about the fact that they must do better by public school students. Rather, the public has a sense that America is failing its students.

- An 80% majority of adults and 84% of high school parents identify as a big problem a belief that students are getting passed through the public high school system without the skills they need for college or work.
- More than three-quarters of the public (76%) and high school administrators (86%) identify students' dropping out of high school as at least a fairly big problem.
- Somewhat fewer, although still a majority (64%), Americans and 84% of college faculty and opinion leaders identify as a big problem a belief that gifted students are not being sufficiently challenged so that they are ready to compete against the best-educated scientists and engineers in the world.

Americans understand that we must take on the difficult challenge of reforming our schools so that they best serve America's youth as they prepare to enter the work force and take on the responsibility of ensuring American global competitiveness, and therefore American security, into the future. The public is willing to make sacrifices and understands that it is time to make substantive changes to the way that classes and curriculum are structured; the way teachers are recruited, hired, and rewarded; and the way funding and resources are allocated to the public school systems.

The public is very receptive to reform proposals. Even when presented with the tradeoff of a significant increase in their own taxes, the public overwhelmingly supports a wide variety of reform proposals, delineated below.

### Raising standards

- increase expectations for parental involvement with their children's education (93% favor)

- challenge and inspire students at high risk of dropping out by increasing resources, lowering class sizes, and raising expectations (88% favor)
- develop more academically rigorous standards for high schools with greater emphasis on college preparatory classes (87% favor)
- increase the number of students pursuing careers in math and science by attracting more math and science teachers through a variety of financial and in-kind incentives, such as loan forgiveness and housing vouchers (85% favor)
- require students to pass statewide graduation tests ensuring they have mastered the core subject areas (81% favor)

### Giving students more options

- place greater emphasis on real-world learning for students, such as work-study programs and vocational training (92% favor)
- expand after-school programs and lower the class size in elementary schools, even if this increases the per-pupil cost by thousands of dollars (81% favor)
- make more college course work and apprenticeships available to high school students, even if this takes away from time students spend on the core curriculum (75% favor)

### Improving teaching

- improve the quality of teachers by ensuring they are experts in the subjects they teach, including providing more training and requiring certification (92% favor)
- overhaul teacher-hiring practices to ensure that passionate and talented teachers are hired; and create a "master teacher" ladder to reward teaching as a career (91% favor)
- increase investments in low-income schools to improve their infrastructure and materials and to recruit and retain the most-gifted teachers, even if

this entails shifting funds from middle- and high-income areas (77% favor)

- dramatically increase teachers' salaries to attract more well-qualified teachers, even if this entails a significant increase in taxes (73% favor)

The public understands that America's future success in the global economy rests upon improving our public education system so that America's youth can compete with students from around the world for highly technical scientific and engineering jobs. Americans believe that it is time for everyone — students, teachers, parents, administrators, and the community as a whole — to dramatically increase our efforts to ensure that students get the most out of their public education. The public is eager to embrace a variety of reforms and is demanding that policymakers and education experts step up and lead by identifying which reform approaches will be most effective at improving our public schools.

### Methodology

From May 22 to June 8, 2006, Peter D. Hart Research Associates, Inc., and The Winston Group conducted a national survey among 1,215 adults, including 512 parents of K-12 students (301 parents of high school students), and among 231 public high school students, 150 high school administrators (superintendents, school board members, principals and vice principals), 150 high school teachers, 151 college faculty, and 151 opinion leaders (persons in business, association/advocacy, and state/local government). At the 95% confidence level, the data's margin of error is  $\pm 3.1$  percentage points among all adults, and larger among the subsamples of parents of K-12 students ( $\pm 4.3$ ), parents of high school students ( $\pm 5.7$ ), high school students ( $\pm 6.5$ ), high school administrators ( $\pm 8.0$ ), high school teachers ( $\pm 8.0$ ), college faculty ( $\pm 8.0$ ), and opinion leaders ( $\pm 8.0$ ).

Prior to conducting the surveys, Hart Research and The Winston Group convened seven focus groups: two groups in Richmond, Va.; two in Encino, Calif.; one in Rosemont, Ill.; and two in Atlanta, Ga. One group survey each was conducted among male high school students, female high school students, parents of middle and/or high school students, high school teachers, high school administrators, college professors, and business leaders/HR professionals. In addition, seven in-depth interviews were conducted among leading competitiveness experts.

### Additional Survey Data

To view the full range of America Speaks surveys, please visit ETS at [www.ets.org](http://www.ets.org), then click on the link titled "Policymakers," followed by the link titled "Americans Speak." At that page, you will find the following survey reports:

- *Ready for the Real World? Americans Speak on High School Reform (June 2005)*
- *Equity and Adequacy: Americans Speak on Public Education Funding (June 2004)*
- *Quality, Affordability and Access: Americans Speak on Higher Education (June 2003)*
- *A National Priority: Americans Speak on Teacher Quality (June 2002)*
- *A Measured Response: Americans Speak on Education Reform (May 2001)*

# Student Performance Data

## ***National Assessment of Educational Progress (NAEP) Scores***

The National Assessment of Educational Progress (NAEP), also known as “the Nation’s Report Card,” is the only nationally representative and continuing assessment of what America’s students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography and the arts.

NAEP does not provide scores for individual students or schools; instead, it offers results regarding subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., fourth-graders) and groups within those populations (e.g., female students, Hispanic students). NAEP results are based on a sample of student populations of interest.

All NAEP data cited is courtesy of the United States Department of Education’s National Center on Education Statistics, and is available on the Internet at the following address:

*[www.nces.ed.gov/nationsreportcard/nde/criteria.asp](http://www.nces.ed.gov/nationsreportcard/nde/criteria.asp)*

## ***2004-05 Adequate Yearly Progress (AYP)/School Improvement Data***

Reprinted with permission from *Education Daily*™. Copyright © 2006 by LRP Publications, PO Box 24668, West Palm Beach, FL 33416-4668. All rights reserved. For more information about *Education Daily*™, please visit *[www.educationdaily.net](http://www.educationdaily.net)* or call 1 (800) 341-7874.

## ***Scholastic Aptitude Test (SAT®) Scores***

*SAT Reasoning Test*™ Verbal and Math Scores by State With Changes for Selected Years, Copyright © 2005 [www.collegeboard.com](http://www.collegeboard.com). Reproduced with permission. All rights reserved.

## AVERAGE NAEP MATHEMATICS SCALE SCORES, GRADE 4 PUBLIC SCHOOLS: BY STATE, 2000-05

YEAR	2000 Accommodations NOT permitted	2000 Accommodations Permitted	2003	2005
National Average	226	224	234	237
Alabama	218	217	223	225
Alaska	-	-	233	236
Arizona	219	219	229	230
Arkansas	217	216	229	236
California	214	213	227	230
Colorado	-	-	235	239
Connecticut	234	234	241	242
Delaware	-	-	236	240
District of Columbia	193	192	205	211
Florida	-	-	234	239
Georgia	220	219	230	234
Hawaii	216	216	227	230
Idaho	227	224	235	242
Illinois	225	223	233	233
Indiana	234	233	238	240
Iowa	233	231	238	240
Kansas	232	232	242	246
Kentucky	221	219	229	231
Louisiana	218	218	226	230
Maine	231	230	238	241
Maryland	222	222	233	238
Massachusetts	235	233	242	247
Michigan	231	229	236	238
Minnesota	235	234	242	246
Mississippi	211	211	223	227
Missouri	229	228	235	235
Montana	230	228	236	241
Nebraska	226	225	236	238
Nevada	220	220	228	230
New Hampshire	-	-	243	246
New Jersey	-	-	239	244
New Mexico	214	213	223	224
New York	227	225	236	238
North Carolina	232	230	242	241
North Dakota	231	230	238	243
Ohio	231	230	238	242
Oklahoma	225	224	229	234
Oregon	227	224	236	238
Pennsylvania	-	-	236	241
Rhode Island	225	224	230	233
South Carolina	220	220	236	238
South Dakota	-	-	237	242
Tennessee	220	220	228	232
Texas	233	231	237	242
Utah	227	227	235	239
Vermont	232	232	242	244
Virginia	230	230	239	240
Washington	-	-	238	242
West Virginia	225	223	231	231
Wisconsin	-	-	237	241
Wyoming	229	229	241	243

## PERCENTAGE OF NAEP STUDENTS AT OR ABOVE PROFICIENT IN MATHEMATICS, GRADE 4 PUBLIC SCHOOLS: BY STATE, 2000-05

YEAR	2000 Accommodations NOT permitted	2000 Accommodations Permitted	2003	2005
National Average	25	22	31	35
Alabama	14	13	19	21
Alaska	-	-	30	34
Arizona	17	16	25	28
Arkansas	13	14	26	34
California	15	13	25	28
Colorado	-	-	34	39
Connecticut	32	31	41	42
Delaware	-	-	31	36
District of Columbia	6	5	7	10
Florida	-	-	31	37
Georgia	18	17	27	30
Hawaii	14	14	23	27
Idaho	21	20	31	40
Illinois	21	20	32	32
Indiana	31	30	35	38
Iowa	28	26	36	37
Kansas	30	29	41	47
Kentucky	17	17	22	26
Louisiana	14	14	21	24
Maine	25	23	34	39
Maryland	22	21	31	38
Massachusetts	33	31	41	49
Michigan	29	28	34	38
Minnesota	34	33	42	47
Mississippi	9	9	17	19
Missouri	23	23	30	31
Montana	25	24	31	38
Nebraska	24	24	34	36
Nevada	16	16	23	26
New Hampshire	-	-	43	47
New Jersey	-	-	39	45
New Mexico	12	12	17	19
New York	22	21	33	36
North Carolina	28	25	41	40
North Dakota	25	25	34	40
Ohio	26	25	36	43
Oklahoma	16	16	23	29
Oregon	23	23	33	37
Pennsylvania	-	-	36	41
Rhode Island	23	22	28	31
South Carolina	18	18	32	36
South Dakota	-	-	34	41
Tennessee	18	18	24	28
Texas	27	25	33	40
Utah	24	23	31	37
Vermont	29	29	42	44
Virginia	25	24	36	39
Washington	-	-	36	42
West Virginia	18	17	24	25
Wisconsin	-	-	35	40
Wyoming	25	25	39	43

## AVERAGE NAEP MATHEMATICS SCALE SCORES, GRADE 8 PUBLIC SCHOOLS: BY STATE, 2000-05

YEAR	2000 Accommodations NOT permitted	2000 Accommodations Permitted	2003	2005
National Average	274	272	276	278
Alabama	262	264	262	262
Alaska	-	-	279	279
Arizona	271	269	271	274
Arkansas	261	257	266	272
California	262	260	267	269
Colorado	-	-	283	281
Connecticut	282	281	284	281
Delaware	-	-	277	281
District of Columbia	234	235	243	245
Florida	-	-	271	274
Georgia	266	265	270	272
Hawaii	263	262	266	266
Idaho	278	277	280	281
Illinois	277	275	277	278
Indiana	283	281	281	282
Iowa	-	-	284	284
Kansas	284	283	284	284
Kentucky	272	270	274	274
Louisiana	259	259	266	268
Maine	284	281	282	281
Maryland	276	272	278	278
Massachusetts	283	279	287	292
Michigan	278	277	276	277
Minnesota	288	287	291	290
Mississippi	254	254	261	262
Missouri	274	271	279	276
Montana	287	285	286	286
Nebraska	281	280	282	284
Nevada	268	265	268	270
New Hampshire	-	-	286	285
New Jersey	-	-	281	284
New Mexico	260	259	263	263
New York	276	271	280	280
North Carolina	280	276	281	282
North Dakota	283	282	287	287
Ohio	283	281	282	283
Oklahoma	272	270	272	271
Oregon	281	280	281	282
Pennsylvania	-	-	279	281
Rhode Island	273	269	272	272
South Carolina	266	265	277	281
South Dakota	-	-	285	287
Tennessee	263	262	268	271
Texas	275	273	277	281
Utah	275	274	281	279
Vermont	283	281	286	287
Virginia	277	275	282	284
Washington	-	-	281	285
West Virginia	271	266	271	269
Wisconsin	-	-	284	285
Wyoming	277	276	284	282

## PERCENTAGE OF NAEP STUDENTS AT OR ABOVE PROFICIENT IN MATHEMATICS, GRADE 8 PUBLIC SCHOOLS: BY STATE, 2000-05

YEAR	2000 Accommodations NOT permitted	2000 Accommodations Permitted	2003	2005
National Average	26	25	27	28
Alabama	16	16	16	15
Alaska	-	-	30	29
Arizona	21	20	21	26
Arkansas	14	13	19	22
California	18	17	22	22
Colorado	-	-	34	32
Connecticut	34	33	35	35
Delaware	-	-	26	30
District of Columbia	6	6	6	7
Florida	-	-	23	26
Georgia	19	19	22	23
Hawaii	16	16	17	18
Idaho	27	26	28	30
Illinois	27	26	29	29
Indiana	31	29	31	30
Iowa	-	-	33	34
Kansas	34	34	34	34
Kentucky	21	20	24	23
Louisiana	12	11	17	16
Maine	32	30	29	30
Maryland	29	27	30	30
Massachusetts	32	30	38	43
Michigan	28	28	28	29
Minnesota	40	39	44	43
Mississippi	8	9	12	14
Missouri	22	21	28	26
Montana	37	36	35	36
Nebraska	31	30	32	35
Nevada	20	18	20	21
New Hampshire	-	-	35	35
New Jersey	-	-	33	36
New Mexico	13	12	15	14
New York	26	24	32	31
North Carolina	30	27	32	32
North Dakota	31	30	36	35
Ohio	31	30	30	33
Oklahoma	19	18	20	21
Oregon	32	31	32	34
Pennsylvania	-	-	30	31
Rhode Island	24	22	24	24
South Carolina	18	17	26	30
South Dakota	-	-	35	36
Tennessee	17	16	21	21
Texas	24	24	25	31
Utah	26	25	31	30
Vermont	32	31	35	38
Virginia	26	25	31	33
Washington	-	-	32	36
West Virginia	18	17	20	18
Wisconsin	-	-	35	36
Wyoming	25	23	32	29

## AVERAGE NAEP READING SCALE SCORES, GRADE 4 PUBLIC SCHOOLS: BY STATE, 1998-2005

YEAR	1998	2002	2003	2005
National Average	213	217	216	217
Alabama	211	207	207	208
Alaska	-	-	212	211
Arizona	206	205	209	207
Arkansas	209	213	214	217
California	202	206	206	207
Colorado	220	-	224	224
Connecticut	230	229	228	226
Delaware	207	224	224	226
District of Columbia	179	191	188	191
Florida	206	214	218	219
Georgia	209	215	214	214
Hawaii	200	208	208	210
Idaho	-	220	218	222
Illinois	-	-	216	216
Indiana	-	222	220	218
Iowa	220	223	223	221
Kansas	221	222	220	220
Kentucky	218	219	219	220
Louisiana	200	207	205	209
Maine	225	225	224	225
Maryland	212	217	219	220
Massachusetts	223	234	228	231
Michigan	216	219	219	218
Minnesota	219	225	223	225
Mississippi	203	203	205	204
Missouri	216	220	222	221
Montana	225	224	223	225
Nebraska	-	222	221	221
Nevada	206	209	207	207
New Hampshire	226	-	228	227
New Jersey	-	-	225	223
New Mexico	205	208	203	207
New York	215	222	222	223
North Carolina	213	222	221	217
North Dakota	-	224	222	225
Ohio	-	222	222	223
Oklahoma	219	213	214	214
Oregon	212	220	218	217
Pennsylvania	-	221	219	223
Rhode Island	218	220	216	216
South Carolina	209	214	215	213
South Dakota	-	-	222	222
Tennessee	212	214	212	214
Texas	214	217	215	219
Utah	216	222	219	221
Vermont	-	227	226	227
Virginia	217	225	223	226
Washington	218	224	221	223
West Virginia	216	219	219	215
Wisconsin	222	-	221	221
Wyoming	218	221	222	223

## PERCENTAGE OF NAEP STUDENTS AT OR ABOVE PROFICIENT IN READING, GRADE 4 PUBLIC SCHOOLS: BY STATE, 1998-2005

YEAR	1998	2002	2003	2005
National Average	28	30	30	
Alabama	24	22	22	22
Alaska	-	-	28	27
Arizona	22	22	23	24
Arkansas	23	26	28	30
California	20	21	21	21
Colorado	33	-	37	37
Connecticut	43	43	43	38
Delaware	22	35	33	34
District of Columbia	10	10	10	11
Florida	22	27	32	30
Georgia	24	28	27	26
Hawaii	17	21	21	23
Idaho	-	32	30	33
Illinois	-	-	31	29
Indiana	-	33	33	30
Iowa	33	35	35	33
Kansas	34	34	33	32
Kentucky	29	30	31	31
Louisiana	17	20	20	20
Maine	35	35	36	35
Maryland	27	30	32	32
Massachusetts	35	47	40	44
Michigan	28	30	32	32
Minnesota	35	37	37	38
Mississippi	17	16	18	18
Missouri	28	32	34	33
Montana	37	36	35	36
Nebraska	-	34	32	34
Nevada	20	21	20	21
New Hampshire	37	-	40	39
New Jersey	-	-	39	37
New Mexico	21	21	19	20
New York	29	35	34	33
North Carolina	27	32	33	29
North Dakota	-	34	32	35
Ohio	-	34	34	34
Oklahoma	30	26	26	25
Oregon	26	31	31	29
Pennsylvania	-	34	33	36
Rhode Island	31	32	29	30
South Carolina	22	26	26	26
South Dakota	-	-	33	33
Tennessee	25	25	26	27
Texas	28	28	27	29
Utah	28	33	32	34
Vermont	-	39	37	39
Virginia	30	37	35	37
Washington	30	35	33	36
West Virginia	28	28	29	26
Wisconsin	34	-	33	33
Wyoming	29	31	34	34

## AVERAGE NAEP READING SCALE SCORES, GRADE 8 PUBLIC SCHOOLS: BY STATE, 1998-2005

YEAR	1998 Accommodations NOT permitted	2002 Accommodations Permitted	2003	2005
National Average	261	263	261	260
Alabama	255	253	253	252
Alaska	-	-	256	259
Arizona	260	257	255	255
Arkansas	256	260	258	258
California	252	250	251	250
Colorado	264	-	268	265
Connecticut	270	267	267	264
Delaware	254	267	265	266
District of Columbia	236	240	239	238
Florida	255	261	257	256
Georgia	257	258	258	257
Hawaii	249	252	251	249
Idaho	-	266	264	264
Illinois	-	-	266	264
Indiana	-	265	265	261
Iowa	-	-	268	267
Kansas	268	269	266	267
Kentucky	262	265	266	264
Louisiana	252	256	253	253
Maine	271	270	268	270
Maryland	261	263	262	261
Massachusetts	269	271	273	274
Michigan	-	265	264	261
Minnesota	265	-	268	268
Mississippi	251	255	255	251
Missouri	262	268	267	265
Montana	271	270	270	269
Nebraska	-	270	266	267
Nevada	258	251	252	253
New Hampshire	-	-	271	270
New Jersey	-	-	268	269
New Mexico	258	254	252	251
New York	265	264	265	265
North Carolina	262	264	262	258
North Dakota	-	268	270	270
Ohio	-	268	267	267
Oklahoma	265	262	262	260
Oregon	266	268	264	263
Pennsylvania	-	265	264	267
Rhode Island	264	262	261	261
South Carolina	255	258	258	257
South Dakota	-	-	270	269
Tennessee	258	260	258	259
Texas	261	262	259	258
Utah	263	263	264	262
Vermont	-	272	271	269
Virginia	266	269	268	268
Washington	264	268	264	265
West Virginia	262	264	260	255
Wisconsin	265	-	266	266
Wyoming	263	265	267	268

## PERCENTAGE OF NAEP STUDENTS AT OR ABOVE PROFICIENT IN READING, GRADE 8 PUBLIC SCHOOLS: BY STATE, 1998-2005

YEAR	1998 Accommodations NOT permitted	2002 Accommodations Permitted	2003	2005
National Average	30	31	30	29
Alabama	22	21	22	22
Alaska	-	-	27	26
Arizona	27	23	25	23
Arkansas	23	27	27	26
California	21	20	22	21
Colorado	30	-	36	32
Connecticut	40	37	37	34
Delaware	23	33	31	30
District of Columbia	11	10	10	12
Florida	23	29	27	25
Georgia	25	26	26	25
Hawaii	19	20	22	18
Idaho	-	34	32	32
Illinois	-	-	35	21
Indiana	-	32	33	28
Iowa	-	-	36	34
Kansas	36	38	35	35
Kentucky	30	32	34	31
Louisiana	17	22	22	20
Maine	41	38	37	38
Maryland	31	32	31	30
Massachusetts	38	39	43	44
Michigan	-	32	32	28
Minnesota	36	-	37	37
Mississippi	19	20	21	18
Missouri	28	33	34	31
Montana	40	37	37	37
Nebraska	-	36	35	35
Nevada	23	19	21	22
New Hampshire	-	-	40	38
New Jersey	-	-	37	38
New Mexico	23	20	20	19
New York	32	32	35	33
North Carolina	30	32	29	27
North Dakota	-	35	38	37
Ohio	-	35	34	36
Oklahoma	30	28	30	25
Oregon	35	37	33	33
Pennsylvania	-	35	32	36
Rhode Island	32	30	30	29
South Carolina	22	24	24	25
South Dakota	-	-	39	35
Tennessee	27	28	26	26
Texas	27	31	26	26
Utah	31	32	32	29
Vermont	-	40	39	37
Virginia	33	37	36	36
Washington	32	37	33	34
West Virginia	28	29	25	22
Wisconsin	34	-	37	35
Wyoming	31	31	34	36

## 2004-05 NATIONAL AYP, SCHOOL IMPROVEMENT DATA

This data on Adequate Yearly Progress and schools and districts identified as needing improvement under the No Child Left Behind Act was compiled primarily from information available on state education department Web sites, but some of it was obtained directly from state officials.

	Total Schools	Fail to Make AYP	% Fail	Title I Schs	# in Improvement	% of total	Incl. Non Title I	Sanctions applied	% Eligible for Sanctions	% of Total	Yr 1 (Choice)	Yr 2 (SES)	Yr 3 (Correc. Action)	Yr 4+ (Restructuring)	# LEAs	Fail to Make AYP	% Fail	LEAs in Improvement	% in Improvement
AK	495	203	41	291	120	24.2	N	120	40	24.2	24	48	34	14	54	36	66.7	31	57.4
AL	1,366	641	46.9	876	470	34.4	Y	311	36.2	22.8	245	24	1	41	131	42	32.1	42	32.1
AR	1,137	NA	NA	NA	274	24.1	N	274	33.3	24.1	74	151	44	5	252	NA	NA	NA	NA
AZ	1,780	236	13.3	NA	149	8.4	N	149	15.2	8.4	56	42	24	27	538	131	24.3		0
CA	9,395	3,651	38.9	5,840	1,748	18.6	N	1,748	30.5	18.6	397	540	407	404	1,035	417	40.3	151	14.6
CO	1,852	458	24.7	NA	105	5.7	N	105	16.7	5.7	36	31	22	16	182	74	40.7	62	34.1
CT	979	196	20	NA	136	13.9	Y	97	21.7	9.9	16	72	4	5	171	30	17.5	27	15.8
DC	253	62	24.5	NA	109	43.1	NA	109	43.1	43.1	74	0	35	0	—	—	—	—	—
DE	181	47	26	94	45	24.9	Y	38	21	21	10	22	4	2	19	6	31.6	0	0
FL	3,105	1,994	64.2	1,386	1,003	32.3	N		70.4	32.3	328	642	33	0	67	67	100	67	100
GA	2,040	370	18.1	1,144	353	17.3	Y	353	17.3	17.3	146	93	29	85	183	100	54.6	12	6.6
HI	282	185	65.6	NA	136	48.2	Y	136	48.2	48.2	15	65	2	54	—	—	—	—	—
IA	1,532	144	9.4	687	94	6.1	Y	12	1.7	0.8	9	3	0	0	367	24	6.5	14	3.8
ID	601	261	43.4	401	88	14.6	Y	88	14.6	14.6	42	46	0	0	118	73	61.9	47	39.8
IL	3,767	992	26.3	2,335	729	19.4	N	729	31.5	19.4	131	116	151	331	879	237	27	240	27.3
IN	1,862	750	40.3	781	99	5.3	N	99	12.5	5.3	41	22	11	25	293	35	11.9	30	10.2
KS	1,400	121	8.6	NA	15	1.1	N	15	2.3	1.1	8	5	0	2	301	20	6.6	7	2.3
KY	1,172	304	25.9	856	123	10.5	N	123	14.3	10.5	54	59	4	6	176	100	56.8	70	39.8
LA	1,348	219	16.2	NA	155	11.5	Y	155	11.5	11.5	101	20	26	8	68	NA	NA	NA	NA
MA	1,690	836	49.5	NA	409	24.2	Y	409	24.2	24.2	158	185	37	29	242	190	78.5	156	64.5
MD	1,372	349	25.4	NA	165	12	Y	165	12	12	51	38	9	67	25	NA	NA	NA	NA
ME	706	207	29.3	NA	48	6.8	Y	23	3.3	3.3	21	2	0	0	236	NA	NA	NA	NA
MI	3,670	436	11.9	NA	489	13.3	N	489	20	13.3	219	122	32	116	542	24	4.4		0
MN	1,975	244	12.4	873	79	4	N	79	8.8	4	55	16	7	1	479	77	16.1	49	10.2
MO	2,035	708	34.8	1,088	125	6.1	N	125	10.2	6.1	12	106	0	7	NA	NA	NA	NA	NA
MS	782	96	12.3	NA	80	10.2	N	80	21.3	10.2	54	24	0	2	152	35	23	36	23.7
MT	848	52	6.1	NA	28	3.3	Y	67	10.7	7.9	24	9	1	33	405	29	7.2	56	13.8
NC	2,300	964	41.9	NA	195	8.5	N	195	16.9	8.5	96	80	13	6	115	108	93.9	43	37.4
ND	486	43	8.8	NA	18	3.7	N	18	5	3.7	2	0	0	16	202	21	10.4	13	6.4
NE	1,196	56	4.7	NA	5	0.4	N	5	1.1	0.4	3	0	2	0	477	28	5.9	2	0.4
NH	464	37	8	NA	77	16.6	Y	28	29.8	16.6	63	13	1	0	164	NA	NA	NA	NA
NJ	2,398	851	35.5	NA	574	23.9	Y	574	42	23.9	167	301	35	71	593	56	9.4	62	10.5
NM	769	538	70	NA	135	17.6	N	135	28.3	17.6	24	34	18	59	89	27	30.3	0	0
NV	608	323	53.1	NA	174	28.6	Y	174	28.6	28.6	74	82	16	2	17	3	17.6	9	52.9
NY	4,499	1,113	24.7	NA	511	11.4	N	511	18.8	11.4	132	85	100	194	730	235	32.2	56	7.7
OH	3,838	932	24.3	NA	501	13.1	Y	501	13.1	13.1	260	159	26	56	609	271	44.5	59	9.7
OK	1,782	55	3.1	NA	122	6.8	Y	98	8.3	5.5	70	18	3	7	540	48	8.9	22	4.1
OR	1,199	389	32.4	578	283	23.6	Y	42	7.3	3.5	26	14	1	1	197	130	66	25	12.7
PA	3,011	700	23.2	NA	298	9.9	N	298	16	9.9	46	133	30	89	500	192	38.4	33	6.6
RI	316	28	8.9	NA	28	8.9	N	28	20.6	8.9	15	6	6	1	36	NA	NA	NA	NA
SC	1,083	575	53.1	518	165	15.2	N	165	29.5	15.2	36	86	28	15	85	68	80	61	71.8
SD	695	112	16.1	NA	103	14.8	Y	59	16.9	8.5	21	23	13	2	165	3	1.8	5	3
TN	1,693	126	7.4	NA	159	9.4	Y	159	17.7	9.4	96	24	14	25	136	8	5.9	26	19.1
TX	7,020	816	11.6	5,026	176	2.5	N	176	3.5	2.5	115	58	3	0	1,229	131	10.7	14	1.1
UT	943	119	12.6	217	17	1.8	N	17	8	1.8	12	3	1	1	60	26	43.3	25	41.7
VA	1,821	338	18.6	766	312	17.1	Y	104	13.7	5.7	64	27	9	4	132	68	51.5	79	59.8
VT	307	10	3.3	NA	23	7.5	Y	23	10.9	7.5	17	4	1	1	66	6	9.1	10	15.2
WA	2,060	404	19.6	NA	180	8.7	Y	66	6.8	3.2	20	31	6	9	296	87	29.4	29	9.8
WI	2,303	49	2.1	NA	45	2	Y	37	3.3	1.6	14	6	10	7	426	1	0.2	1	0.2
WV	715	121	16.9	NA	36	5	N	36	8.4	5	21	13	1	1	51	13	25.5	28	54.9
WY	362	74	20.4	NA	13	3.6	Y	5	3	1.4	1	3	0	0	48	4	8.3	0	0
US	89,493	22,868	25.6	NA	11,524	12.9	—	10,555	19	11.8	3,757	3,696	1,254	1,847	13,878*	3,281*	23.7*	1,712*	12.4*

## MEAN SAT REASONING TEST™ VERBAL AND MATH SCORES BY STATE, WITH CHANGES FOR SELECTED YEARS

(States are listed by percent of high school graduates who took the SAT Reasoning Test\*.)

The College Board strongly discourages the comparison or ranking of states on the basis of SAT scores alone.

STATE	% Participation Rate 2005*	2005		2004		One-Year Change		2000		Five-Year Change		1995		Ten-Year Change	
		Verbal Mean	Math Mean	Verbal Mean	Math Mean	Verbal	Math	Verbal Mean	Math Mean	Verbal	Math	Verbal Mean	Math Mean	Verbal	Math
New York	92	497	511	497	510	0	1	494	506	3	5	495	498	2	13
Connecticut	86	517	517	515	515	2	2	508	509	9	8	50	502	10	15
Massachusetts	86	520	527	518	523	2	4	511	513	9	14	505	502	15	25
New Jersey	86	503	517	501	514	2	3	498	513	5	4	496	503	7	14
New Hampshire	81	525	525	522	521	3	4	520	519	5	6	520	515	5	10
District of Columbia**	79	490	478	489	476	1	2	494	486	-4	-8	485	471	5	7
Georgia	75	497	496	494	493	3	3	488	486	9	10	483	477	14	19
Maine	75	509	505	505	501	4	4	504	500	5	5	504	497	5	8
Pennsylvania	75	501	503	501	502	0	1	498	497	3	6	496	489	5	14
Delaware	74	503	502	500	499	3	3	502	496	1	6	505	494	-2	8
North Carolina	74	499	511	499	507	0	4	492	496	7	15	488	482	11	29
Virginia	73	516	514	515	509	1	5	509	500	7	14	504	494	12	20
Rhode Island	72	503	505	503	502	0	3	505	500	-2	5	502	490	1	15
Maryland	71	511	515	511	515	0	0	507	509	4	6	506	503	5	12
Vermont	67	521	517	516	512	5	5	513	508	8	9	506	499	15	18
Indiana	66	504	508	501	506	3	2	498	501	6	7	492	494	12	14
Florida	65	498	498	499	499	-1	-1	498	500	0	-2	497	496	1	2
South Carolina	64	494	499	491	495	3	4	484	482	10	17	478	473	16	26
Hawaii	61	490	516	487	514	3	2	488	519	2	-3	483	507	7	9
Oregon	59	526	528	527	528	-1	0	527	527	-1	1	525	522	1	6
Washington	55	532	534	528	531	4	3	526	528	6	6	519	517	13	17
Texas	54	493	502	493	499	0	3	493	500	0	2	495	501	-2	1
Alaska	52	523	519	518	514	5	5	519	515	4	4	521	513	2	6
California	50	504	522	501	519	3	3	497	518	7	4	492	509	12	13
Nevada	39	508	513	507	514	1	-1	510	517	-2	-4	511	508	-3	5
Arizona	33	526	530	523	524	3	6	521	523	5	7	524	520	2	10
Montana	31	540	540	537	539	3	1	543	546	-3	-6	549	553	-9	-13
Ohio	29	539	543	538	542	1	1	533	539	6	4	536	535	3	8
Colorado	26	560	560	554	553	6	7	534	537	26	23	538	532	22	22
Idaho	21	544	542	540	539	4	3	540	541	4	1	544	532	0	10
West Virginia	20	523	511	524	514	-1	-3	526	511	-3	0	525	509	-2	2
Tennessee	16	572	563	567	557	5	6	563	553	9	10	571	560	1	3
New Mexico	13	558	547	554	543	4	4	549	543	9	4	559	549	-1	-2
Kentucky	12	561	559	559	557	2	2	548	550	13	9	552	542	9	17
Wyoming	12	544	543	551	546	-7	-3	545	545	-1	-2	551	544	-7	-1
Minnesota	11	592	597	587	593	5	4	581	594	11	3	580	591	12	6
Alabama	10	567	559	560	553	7	6	559	555	8	4	565	555	2	4
Illinois	10	594	606	585	597	9	9	568	586	26	20	563	574	31	32
Michigan	10	568	579	563	573	5	6	557	569	11	10	559	565	9	14
Kansas	9	585	588	584	585	1	3	574	580	11	8	576	571	9	17
Louisiana	8	565	562	564	561	1	1	562	558	3	4	560	552	5	10
Nebraska	8	574	579	569	576	5	3	560	571	14	8	568	570	6	9
Missouri	7	588	588	587	585	1	3	572	577	16	11	569	566	19	22
Oklahoma	7	570	563	569	566	1	-3	563	560	7	3	565	553	5	10
Utah	7	566	557	565	556	1	1	570	569	-4	-12	585	576	-19	-19
Arkansas	6	563	552	569	555	-6	-3	563	554	0	-2	556	542	7	10
Wisconsin	6	592	599	587	596	5	3	584	597	8	2	574	585	18	14
Iowa	5	596	608	593	602	3	6	589	600	7	8	589	595	7	13
South Dakota	5	589	589	594	597	-5	-8	587	588	2	1	579	576	10	13
Mississippi	4	564	554	562	547	2	7	562	549	2	5	572	557	-8	-3
North Dakota	4	590	605	582	601	8	4	588	609	2	-4	587	602	3	3
All Students	49	508	520	508	518	0	2	505	514	3	6	504	506	4	14

\* Based on the projection of high school graduates in 2005 by the Western Interstate Commission for Higher Education (WICHE), and the number of students in the class of 2005 who took the SAT Reasoning Test.

\*\* 2002 self-reported 12th-grade enrollment from D.C.'s public and nonpublic schools was used since WICHE estimated fewer graduating seniors than actual SAT Reasoning Test takers.

ETS developed this briefing book to provide policy-makers, politicians, thought leaders and citizens with information on key issues in the education-reform debate. It was compiled with the full appreciation of the challenges facing officials responsible for the nation's education system. The opportunities and obstacles involved in reforming our educational system are many and varied.

This document was compiled by Neil Ruddock with the help of several colleagues in ETS's Communications & Public Affairs Division. ETS would like to thank *Education Week*, the College Board, and the U.S. Department of Education's National Council on Education Statistics. Those organizations provided invaluable data that has been included in this publication, and ETS looks forward to working with them to improve education at the federal, state and local levels.

For more information, visit ETS online at

**[www.ets.org](http://www.ets.org)**, or contact

Chuck Melley

Director, Government Relations

ETS

1800 K Street, NW

Suite 900

Washington, DC 20005

[cmelley@ets.org](mailto:cmelley@ets.org)



*Listening. Learning. Leading.*

*[www.ets.org](http://www.ets.org)*

Copyright © 2006 by Educational Testing Service. All rights reserved.  
ETS, the ETS logo, PATHWISE, PRAXIS I, PRAXIS II, PRAXIS III and THE PRAXIS SERIES: PROFESSIONAL ASSESSMENTS FOR BEGINNING TEACHERS are registered trademarks of Educational Testing Service (ETS). PRAXIS and THE PRAXIS SERIES are trademarks of ETS.

SAT is a registered trademark of the College Board.  
SAT REASONING TEST is a trademark of The College Board.  
EDUCATION DAILY is a trademark of LRP Publications