

**Value-Added Assessment:**  
*Powerful Diagnostics to Improve Instruction and Promote Student Achievement*

**Ted Hershberg**  
**Director, Operation Public Education**  
**Professor, Public Policy and History**  
**University of Pennsylvania**

**American Association of School Administrators**  
**Women Administrators Conference 2004 Monograph,**  
**(Forthcoming March 2005, “Leadership in a Time of Change”)**

**Value-Added Assessment: Powerful Diagnostics to Improve Instruction and  
Promote Student Achievement  
Leadership in a Time of Change  
AASA  
March 2005**

*If an unfriendly foreign power had attempted to impose upon America the mediocre educational performance that exists today, we might well have viewed it as an act of war.*

*A Nation at Risk, 1983*

This memorable line from the fire-bell-in-the-night report by President Reagan's National Commission on Excellence in Education for the first time focused attention on the quality of American public education and catalyzed intense reform efforts in our nation's schools. Now more than two decades later we need to ask: "How did we do? What is the state of our schools?"

### **The State of Our Schools**

The latest data from the *National Assessment of Educational Progress* (NAEP) — perhaps best known as "the Nation's Report Card" — make clear that the proportion of our students who score at the proficient level or higher in math, science and reading ranges between 28 and 31 percent. Since fewer than 30 percent of Americans live in big cities, these data should silence those who claim "it's the urban tail that's wagging the school reform dog." In other words, the belief that the suburban schools, where the large majority of our nation's children attend classes, are fine and that the problems are mainly found in urban schools, with their large low-income and minority populations, is unsupportable in fact.

From where, then, does the ill-founded complacency come? The answer may be found by comparing the NAEP scores with the results from tests given by the individual states. The pattern for the vast majority of states shows much higher proportions of students performing at proficient or above on their own tests. In Virginia, for example, in 8<sup>th</sup> grade math, 71 percent of the students score at these levels on the state's test, but only 31 percent do so on the NAEP. In North Carolina in 4<sup>th</sup> grade math the numbers are 90 percent on the state's test, but only 41 percent on the NAEP.

Where to draw the "cut" score — the line that separates "proficient" from "basic" — is not without some controversy. While the NAEP lines are drawn by a panel of some of the nation's most prominent assessment experts, the lines drawn on state tests are strongly influenced by a political question of real interest to legislators: How much failure will our public tolerate? This problem is real and it exists nationwide.

There is neither time nor space here to review in detail all the relevant test data, but the following examples are typical. The *Trends in International Math and Science Survey* (TIMSS, 1995 and 1999) documented that in comparisons among 21 developed nations, American students rank near the top in fourth grade, but fall to the bottom 10 percent by twelfth grade. The reasons for the poor performance of American students have little to do with the number of school days per year (180 in the U.S. vs. 220 or 240 in Western Europe or the Far East) because doing more of the wrong thing doesn't improve test scores. TIMSS concluded that the

responsibility rests with the math and science curriculum taught in America: it was described as “incoherent . . . a mile wide and inch deep.” We teach 25 to 30 fundamental concepts while the other nations focus on five to seven. As a result, our students leave school with a superficial grasp of the key criteria for success in these critical disciplines.

The *New Standards Reference Exams*, the product of the National New Standards Project and published by Harcourt Brace, are widely considered to be the most demanding assessment instruments on the market because they contain mostly open-ended, problem-solving questions that get at higher-order thinking skills. Results from tests administered to a 20 percent sample of fourth, eighth and tenth graders in affluent suburban communities in a major metropolitan area in the Northeast revealed that only one student in four was performing at the international standard in math problem-solving. Unless one subscribes to the notion that American students are genetically inferior to those elsewhere in the developed world, the conclusion is inescapable that our schools are not teaching this skill.

This understanding is corroborated in a spate of studies that attempt to quantify how time is spent in classrooms. Although variations exist by subject and schooling level (lower, middle and high school), it is fair to generalize that three-quarters of classroom time is devoted to lecture; the remainder consists of “Q & A,” but virtually all of this is characterized by “yes” and “no” drills associated with identifying the “right” answer. There is almost a disciplined effort to avoid challenging questions that require thoughtful answers.

A final piece of evidence comes from the Report of the National Research Council on advanced placement released in 2003. Their recommendation was no longer to give college credit for Advanced Placement (AP) courses *unless* the student scored a “5” on the AP tests – the top score possible. According to the report, when students scored below this level they entered college filled with facts about math and science, but they could not reason in scientific and mathematical terms at the level deemed appropriate for higher education. When one acknowledges that AP courses are taught in our best high schools and taken by our best students, it is clear that the memorization paradigm reigns supreme in K12, despite how much Dewey and Piaget students may read in graduate school.

We will consider what the schools of the 21<sup>st</sup> century will need to look like a little later, but for now it is important to understand the “push back” I get when I share these test results data with educators around the nation.

### **The Economic Argument**

“Isn’t the American economy,” I am repeatedly asked, “the envy of the world?” “If so,” the argument continues, “then our schools must be doing an awful lot of things right. Test scores don’t matter — they don’t predict success in later life. Think of our schools as a ‘black box.’ We may not know precisely what goes on inside, but the subsequent success of our graduates is undeniable.”

The fatal flaw with this argument is that it confuses the success of the American economy in the *aggregate* with the success of individual American families. Here the data are unequivocal. Virtually all the income gains of the last 30 years have gone to the top fifth of our families, in large measure according to economists, because of new technologies that favor the better educated.

The literature on the changing skill demands of the new economy is vast and growing. According to a survey done by the American Management Association in 2000, the proportion of workers who lacked skills for entry-level jobs doubled between 1996 and 1999 from 19 percent to 38 percent. The explanation was not that the schools were “dumbing-down” their curricula; rather, the economy was undergoing remarkably rapid change. These were the years when the use of email and the Internet was reaching into businesses, large and small across all sectors.

The fact is, our schools are not getting worse. Despite all the “school-bashing,” the data from the mid-seventies through the mid-nineties suggest that the schools actually improved a bit. But marginal improvement does not mean that our schools are good enough for the 21st century.

Unfortunately, they are not. “Our nation . . . now faces a stark and sudden choice: adapt or perish,” Louis Gerstner, former Chairman of IBM and chairman of The Teaching Commission, wrote recently in the Wall Street Journal. He referred not to the war against terrorism “but to a war of skills” that we’re at risk “of losing to India, China, and other emerging economies . . . We’re losing it not on factory floors or lab benches...but in our public schools. Unless we transform those schools and do it now . . . it will soon be too late.”

The key point here is that no one is to blame — not the teachers’ unions, the administrators, the school boards nor those responsible for funding. The simple reality is that the global economy has changed in fundamental ways and done so far more rapidly than our schools have been able to adapt. It is now critical to move our schools to their next level of excellence.

### **American Public Education: Past and Future**

The current system of public education is largely unchanged from what was put in place in the 19<sup>th</sup> century when it was designed to do three things. The first was to provide universal basic literacy, and America became the first nation to achieve this goal. Virtually everyone in the workforce could read and write at the sixth grade level. The second was to socialize a highly diverse population — millions of immigrants from different nations, cultures and religions and even more millions of farmers who migrated to cities as agriculture mechanized over 150 years — for success in an industrial economy. Schools taught students to show up on time, respect authority, develop a work ethic and repeat monotonous tasks. Finally, using standardized tests (which are the first or second cousins of IQ tests) and the bell-shaped curve, the schools identified and sorted out the top fifth of their students for higher education and the best and brightest among these went on to run the country.

Our schools were enormously successful in these tasks, and there is no way to understand the emergence of America as an industrial superpower in the 20<sup>th</sup> century without acknowledging the key role they played. Our schools and the factory economy were perfectly aligned.

The old system was about quantity and cohorts. It worked like this: we opened the doors in September; in came the kids; they were given the curriculum and tested at the end of the school year. Some kids “got it.” For those who didn’t — live the consequences, next group please. Or in the words of my mother, “So you make your bed, so you lie in it.”

The new system must be about quality and the education of individual students. Instead of holding constant the time and allowing the results to vary, we must hold constant the results (the standards) and vary the time and the resources it takes to help every child achieve them. This “flipping” of the system is at the heart of standards movement, but thus far all we have done is ended the practice of “social promotion” while preserving the status quo organization of our schools.

In the old system the top fifth “got it” and moved on. For the remaining 80 percent, there was little consequence because for most of our history they ventured forth into an economy that provided ubiquitous jobs paying middle-class sustaining wages but required little in the way of education or advanced skills.

That world no longer exists. The last year the typical blue-collar worker earned enough for mom to stay at home and raise the kids was 1964. We maintained our standard of living despite the decline of real wages through these decades largely because women entered the labor force in record numbers creating two-income families, but that strategy has run its course. To preserve the stable, middle-class basis of our society, schools must now graduate their students able to use technology, think critically, solve problems and learn on their own throughout their lives. In sum, we are asking our schools to do two things: educate *all* the children, not simply the top fifth, and educate them to unprecedentedly high levels.

This has never been done before. Worse, the record of “reform” is deeply disappointing. Since 1970, despite an increase in real spending of over 100 percent per pupil, a decrease in the pupil-to-teacher ratio of 22 percent and a doubling of the number of teachers with Masters’ degrees, there has been precious little gain in student achievement.

Why, as we face these twin challenges, should a reasonable person believe that a school system designed for a previous century and a different economy is the right one? The definition of insanity, Albert Einstein was reported to say, is doing the same thing over and over again while expecting different results.

In a new school system — one able to meet the educational challenges of the 21st century — teachers and administrators will have to master new concepts and skills. They will need to replace the “ability-based” notion that dominates the thinking of too many educators and parents with an *effort-based* theory of learning: you are not simply born smart, but you can “get smart” with appropriate resources and high-quality instruction. They will need to see how teaching a *standards-driven curriculum* is very unlike using a bell-shaped curve to distinguish among students: all students must reach high standards. They will need to be proficient in using a *problem-solving pedagogy* in the classroom: memorization will always have a place in the learning process because no subject can be mastered without it, but it must never again serve as the dominant paradigm. They will have to learn how to *differentiate instruction*: “one-size-fits-all” cannot suffice in an era in which no child can be left behind. They must master *data-driven decision making*, a striking departure from an anecdotal approach. Finally, they must develop an ability to create *student-centered classrooms*: lectures serve the needs of some students, but everyone learns best when they bear more of the responsibility for learning.

To support the mastery of these new concept and skills, we will need a dramatic change in the rules and incentives that govern K12. Operation Public Education, a foundation-funded effort based at the University of Pennsylvania, has developed a set of comprehensive reforms that should significantly increase the chances that our schools will succeed in these new tasks. Appendix I provides a brief description of the new system, but today I can focus only on its foundation — a methodology known as *value-added assessment*.

### **Value-Added Assessment: The Cornerstone of Reform**

Our last four presidents, the Congress, governors and corporate leaders came to understand that if America is to remain a stable, middle-class society, steps had to be taken to significantly improve our system of public schools. Frustrated by precious little improvement in student achievement over the last two decades the Congress in a bipartisan consensus in 2002 passed the *No Child Left Behind Act* (NCLB), an unprecedented expansion of the role of the federal government in American K12 education.

NCLB broke new ground. It required each state to establish academic standards and to ensure that all their students met proficiency by 2014. To measure progress toward this goal, states had to meet Adequate Yearly Progress (AYP) targets defined as a fixed percentage of students achieving proficiency. In a further major change, proficiency could not be achieved based on school-wide averages; each of several sub-groups — low-income, non-English speakers, those with special needs, and various racial or ethnic groups — had to meet these targets or the school would be sanctioned. For schools failing to make AYP, warnings would be issued in the first two years; without improvement schools would have to cover the cost of transportation so failing students could attend other schools (year three) and the cost of tutoring (year four); a fifth year without improvement could lead to the closing of the school or its reconstitution.

In its current form, NCLB is unlikely to lead to the necessary changes described above. But if states adopt a new statistical methodology known as value-added assessment that isolates the impact of instruction on student learning, they will provide educators with powerful diagnostic data to improve teaching and put in place the necessary foundation to support a transformation of our schools.

Although there are several forms of value-added assessment, the only one adopted for statewide use was developed by Dr. William Sanders, when he was a professor of statistics at the University of Tennessee. It has been used in that state since 1992. Tennessee now has the largest data base of its type in the world: over 10 million records that include the scores on Terra Nova tests in math, science, reading, writing and social studies for each student and the name of the teacher who taught them. In addition to Tennessee, the Sander's model is currently used in over 300 school districts in 21 states and is being rolled-out for use in all school districts in Pennsylvania and Ohio.

The value-added approach to assessment centers on a disarmingly simple but profound notion: schools cannot solve all of society's problems, but *they can and should ensure that every child receives a year's worth of growth in a year*. A year's worth of growth — whether children start the year below, on, or above grade — is the amount that should be reasonably expected of them based on what they actually achieved in past years. This belief — that each child is entitled to at least this much annual growth — lies at the heart of value-added methodology.

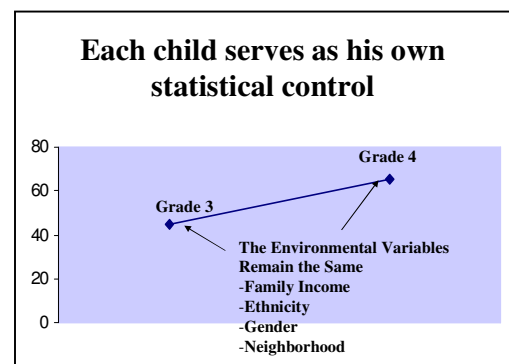
The easiest way to begin understanding value-added assessment is to ask, “What is a good school?” Most people would answer quickly: “It’s a school with high test scores.” The funny thing about high-test scores, however, is that they are invariably found in the schools in wealthy communities. Indeed, according to Educational Testing Service, SAT scores rise with every \$10,000 of family income. This should not be surprising since all the variables that contribute to high-test scores correlate strongly with family income: good jobs, years of schooling, positive attitudes about education, the capacity to expose one’s children to books and travel, and the development of considerable social and intellectual capital that wealthy students bring with them when they enter school. When people understand this they quickly reject the notion that a school is good because its students come from affluent families.

Which leads us the next important question: *Do good schools make good students or do good students make good schools?* One way of penetrating this conundrum would be to undertake a field experiment in which students from wealthy suburban communities would be educated in inner-city schools and students from inner-city communities would be educated in the suburban schools. We would then watch what would happen over time. Because it is not possible to conduct such an experiment, we need a statistical technique that could approximate the results for us. That is what value-added assessment can do: it separates the *student* effects (e.g., ethnicity, family, SES) from the *school* effects (e.g., teachers, administrators and academic programs).

Value-added is not a test — no more classroom time is spent assessing students — but a new way of looking at the results that come from tests so we can determine whether the students in a classroom, school or district are making sufficient academic growth each year.

Value-added assessment is often confused with simple growth because the words themselves make it is easy to think about this growth as the “value” that is “added” over the last year. But the statistical method known as “value-added assessment,” is a way of isolating the impact of instruction on student learning. Its great advantage is its ability to separate the annual academic growth of students into two parts: that which can be attributed to the student and that which can be attributed to the classroom, school or district. Because individual students rather than cohorts are traced over time, each student serves as his or her own “baseline” or control, which removes virtually all of the influence of the unvarying characteristics of the student, such as race or socioeconomic factors.

While value-added is statistically and computationally complex, it is relatively easy to grasp at the conceptual level. Test scores are projected for students and then compared to the scores they actually achieve at the end of the school year. Classroom scores that exceed projected values indicate effective instruction. Conversely, scores that are mostly below projections suggest that the instruction was ineffective.

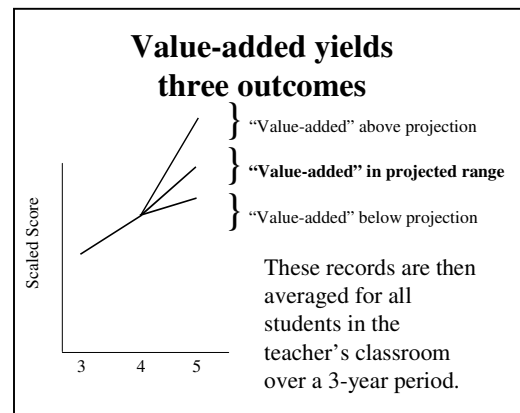


At the same time this approach considers student factors such as the pattern of prior test scores, both those of the individual student as well as those of other students in the same class. If a student’s present performance is below projected scores, while students with comparable pre-

vious academic history in the same classes have done well, this is evidence of the student effect — external variables such as the home environment which are outside the control of teachers and schools.

Value-added is fair to students because their projected scores are based only on their prior academic record. Since it does not consider the student's race or socioeconomic background, no bias is introduced — low-income children are not expected to do poorly and high-income students are not expected to do well. But it is also fair to educators precisely because by using student's past performance value-added takes into account family and neighborhood characteristics that so strongly influence absolute test scores.

There are essentially three outcomes in value-added: classrooms, schools or districts where on average students are receiving a year's worth of growth in year; those in which the students are being "stretched" so that (in a statistically significant sense) they grow more than a year's worth of growth annually and those where they are losing ground meaning their achievement levels fell below where they had demonstrated a capacity to perform in past years. Classroom scores are based on all the students in that classroom for three years; scores for schools and districts are simply aggregations of classrooms and all are expressed as three year-running averages.

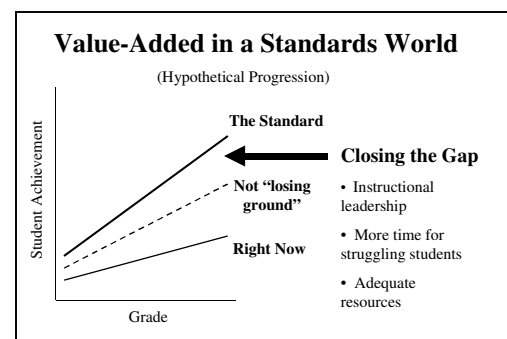


The value-added calculation is concerned not with the score on an achievement test by itself, but with the difference between this actual score and the projected score. Because the key measurement is between these two rather than on the absolute score alone, it does not matter what the mix of students is in a teacher's classroom or in the school or district as a whole. In this sense, value-added levels the playing field across schools of very different socioeconomic levels.

The question being asked, in other words, is not simply how high was the level of achievement (which is always influenced by family income), but how much value did the school staff add to the students who live in its community? How, in effect, did they do with the hand society dealt them?

We can now answer the question, "What is a good school?" in a much more accurate fashion: a school is good if each year the performance of its students exceeds what was expected of them based on past academic achievement. And because NCLB requires all schools to bring all their students to proficiency (by 2014), a good school must also be one that meets this requirement.

Because value-added is a norms-driven approach and schools must now operate in a criterion-referenced or standards environment, the question emerges whether



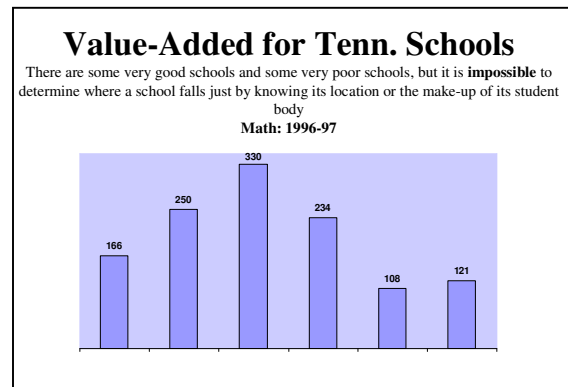
these can co-exist in the same system. The nightmare many teachers face is how to close the gap between the high standards and student performance levels that are far below them.

To close the gap, schools must do two things: not lose ground and make strategic investments. The problem is, American schools lose a great deal of ground every year. So if they can keep their students moving forward by giving them a year's worth of growth in a year, a substantial part of the gap can be closed. The balance must come from new investments: adequate resources and high-quality instructional leadership for the school or district. Superintendents and principals must create an environment that provides their faculties with the requisite resources and professional development to help them help their students perform at high levels; and teachers must provide their students with high-quality instruction and have adequate time to work with struggling students.

### What Value-Added Findings Reveal

Let's turn now to consider some actual data from Tennessee. In the chart below, all 1209 K8 schools in Tennessee are grouped according to the value-added scores they achieved in math in 1996-97. The key number to bear in mind in this example and in several more that follow is "100" — that is, if a school has a score of 100, it means that on average all their students enjoyed a year's worth of growth in a year. Above 100 means that they've "stretched" their students beyond what was projected for them based on past academic achievement, while below 100 means their students have lost ground.

There were 166 schools with scores exceeding 110 percent: these were the highest performing schools in the state, providing their students with far more than year's worth of growth in a year. At the other end of the continuum there were 121 schools with scores between 60 and 70 percent. These were the worst performing schools in the state: their students were falling far short of the growth they had demonstrated a capacity to achieve at in the past.

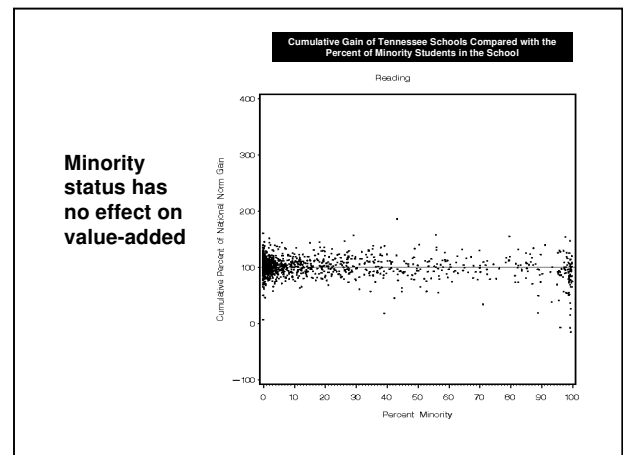
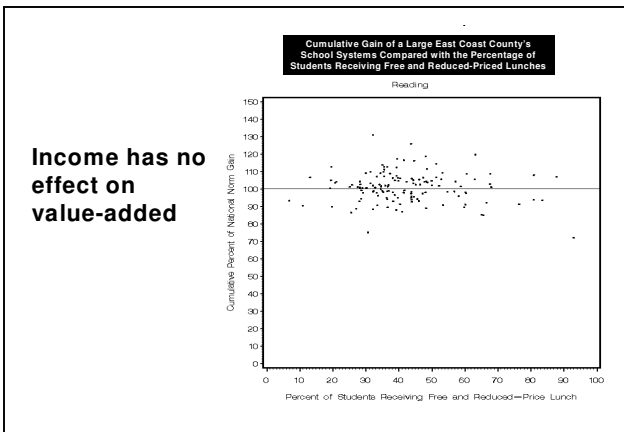


If you assume that an affluent school district with high-test scores would appear in the former group, you may or may not be correct. In fact, in Tennessee many such schools were found in the group with value-added scores of between 90 and 100 percent — meaning they actually lost ground over the year. Sanders labeled these "slide and glide" schools because although they had high-test scores, they were apparently resting on the laurels of their students, thus depriving them of the growth to which they were entitled. When the superintendents of these schools complained that they could not do any better with their already high-scoring students, Sanders pointed to other affluent communities whose students had high achievement scores and whose schools also had high value-added scores.

Even more significant was the fact that many in the highest value-added category were low-income, high minority schools that were doing an excellent job of "stretching" their students, but which never appeared as successful in schools rankings based on achievement.

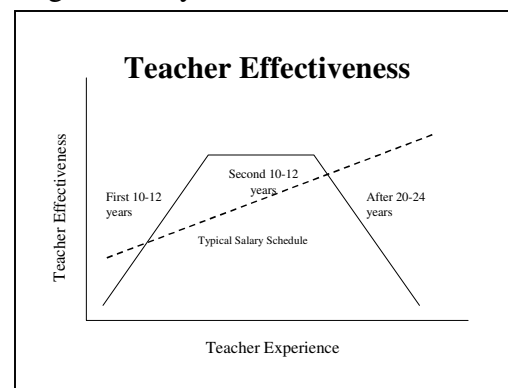
Most striking in the Tennessee data, however, is the fact that 68 percent of the K8 schools lost ground in math in 1996-97. This lost ground is what lurks behind the results of the Trends in International Math and Science Study reported earlier — that American students rank near the top in fourth grade, but are in the bottom 10 percent of the developed world in twelfth grade.

The following two scatter plots demonstrate that value-added is not affected by income or minority status because these variables are accounted for in the methodology that uses each child as his or her own “control.” On the left, results on a nationally-normed reading test for 130 schools in a large East Coast county are arrayed on the vertical axis by their value-added scores and on the horizontal axis by their proportion of their students eligible for “free and reduced price lunch.” As we move from the richest schools to the poorest schools, no relationship can be found between income and value-added; as many schools are adding value as losing value all along the continuum.



In the example on the right, we've arrayed all 1,209 schools in Tennessee along a similar continuum, but have substituted “minority status” for “free and reduced price lunch.” Once again, as you move from the all-white schools to increasingly high-minority schools, no relationship is observed with value-added: as many schools add value as lose value all along the continuum until you reach the end, where there are all minority schools. Here, two striking facts emerge: many of these schools (understand they are all quite low-income) are adding considerable value, but there are many more schools that are losing value. Sanders argues this is not a contradiction of his method, as the next example makes powerfully clear.

When all teachers in Tennessee are arrayed according to their years of service and their value-added scores, three patterns emerge that correspond roughly to each third of their careers. Two things are notable about the beginning of the teachers' career. First, many new teachers add little value when they first enter their classrooms. The low value-added scores observed in the all-minority schools, Sanders argues, result from the inexperience of their teachers rather than the shortcoming of their students. This is an indictment both of the poor job colleges of education do in their teacher preparation programs and of the inadequate job school districts do in giving new teachers



the help they need to succeed in their classrooms. We need to improve pre-service education and provide new teachers with multi-year mentoring programs in-service. Second, the good news is that the learning curve for new teachers increases sharply over the first third of their careers.

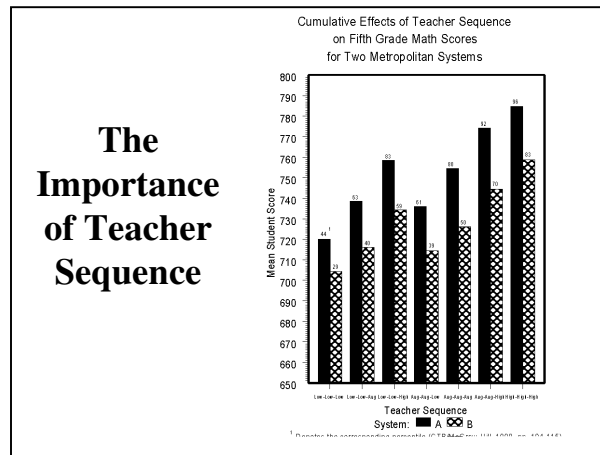
The second third of the career is characterized by a plateau in value-added scores. Teachers maintain their contributions, neither increasing nor decreasing much throughout these years.

After about 20-24 years of service, however, there is a marked decline in teachers' value-added scores. This is undoubtedly due to several factors, including a pattern of attrition that finds more effective teachers leaving the classroom for higher-paying administrative posts or leaving the profession altogether, but most likely it suggests considerable burn-out. When superintendents and principals see this pattern, they often remark that we should mandate retirement after 20 years of service. The problem with that suggestion is that while the variation around the mean in the first two-thirds of a career is relatively small, in the last third it is quite large, meaning that early retirement would force some of the most effective (along with some of the least effective) teachers out of the classroom.

If one overlays the typical salary schedule over these three patterns, it is immediately obvious that schools are paying teachers ever-increasing salaries based on longevity while their average productivity is declining sharply. This system of compensation is simply indefensible now that we can empirically document the profession's productivity levels. If this knowledge were widely known, the tax-paying public would not tolerate it.

Let us turn now to some fascinating studies done in Tennessee using its unique value-added data files that demonstrate the impact of instruction on student learning. In one study of two large metropolitan districts with hundreds of classrooms, researchers set out to examine how the quality of the teacher sequence affected student performance. All students who began third grade at the same level of math achievement were followed through fifth grade and their scale scores on the Terra Nova math test recorded at the end of fifth grade.

The researchers found very different achievement levels for these students depending on the quality of the teachers, as measured by value-added, who taught them math in third, fourth and fifth grades. As expected, students in the wealthier of the two districts (system A) always had higher absolute test scores than their counterparts, but there were striking differences that were attributable to the teacher sequence. Students in both districts with three consecutive teachers drawn from the top third of the teacher pool (ranked by their value-added scores) scored 50 percentile points higher than those who had teachers drawn from the bottom third of teacher pool. In other words, students who started at the same level of math achievement had very different outcomes depending on the quality of the teachers who instructed them in math.



The Dallas Independent School District replicated this study and found the same 50-percentile point difference accounted for by the quality of the teacher sequence. The fact that a dif-

ferent test and a different value-added model were used, but produced the same results — the hallmark of scientific rigor — has persuaded even skeptics that the teacher effect is real and statistically demonstrable.

In a second ground-breaking study, June Rivers, associate director of EVAAS®, demonstrated the powerful impact that instruction has on student learning. Rivers ranked all students in these same large metropolitan Tennessee school districts at the end of fourth grade, dividing them into quartiles. She then asked what the probabilities were for the typical student in each quartile to pass the high stakes exam required for graduation that is given for the first time in ninth grade. In the example I am using here the results are for the bottom quartile, but the patterns she found were the same in each quartile.

Since each student had four teachers in fifth through eighth grades before taking the exam, Rivers identified all teachers in these grades and divided them into quartiles as well, based on their value-added scores. If a typical bottom quartile student had four teachers drawn from the bottom 25 percent of the teacher pool, the chances of passing the test were less than 15 percent.

Typically, explanations for this result would point to factors beyond the school. Some observers cite difficult socioeconomic circumstances. Others suggest inherent ability: some children have it, others don't. The excuses usually conclude with a sad inevitability: "It's most unfortunate, but these are the 'bottom of the barrel' kids." "You can bring the horse to water," we are told, "but you can't make it drink." These explanations share a deeply flawed conclusion: the fault lies with the students rather than with the quality of instruction they received.

Yet, if these same students had four teachers drawn from the middle 50 percent of the teacher pool, their chances of passing increased to 38 percent. And if they were fortunate enough to have had four teachers drawn not from the exceptional top one or five percent of the teacher pool — but from the top 25 percent — their chances of passing the test improved to 60 percent.

Studies like these, based on irrefutable data rather than anecdote, demonstrate the enormous power of effective teaching. Students otherwise deemed as hopelessly low-achievers or academic failures can perform at much higher levels — if they receive high quality instruction. These results also make it difficult to defend a compensation system that pays all teachers the same regardless of their effectiveness with students.

The most successful school in Tennessee by value-added rankings is the Maryville Middle School in suburban Knoxville. Joel Giffin, Maryville's long-time principal, provided data illustrating value-added's cumulative impact. Let's suppose that your company relocates to this area and you decide to move there to keep your job. The purchase of your home, of course, will dictate what school your child will attend. If the home you bought feeds your child into a school with an average value-added score of 100 — students there on average receive a year's worth of growth in a year — your child will graduate eighth grade on grade, exactly what you would expect. But if the home you bought fed your child into a school averaging 75 percent annual growth, he would graduate eighth grade with only 6.5 years of academic

<b>Cumulative Effects of Value-Added</b>			
<b>TVAAS Scores</b>	<b>75%</b>	<b>100%</b>	<b>140%</b>
<b>Grade</b>			
2	2	2	2
3	2.75	3	3.4
4	3.5	4	4.8
5	4.25	5	6.2
6	5	6	7.6
7	5.75	7	9
8	6.5	8	10.4
<b>GRADE LEVEL IMPACT</b>	<b>-1.5</b>	<b>0</b>	<b>+2.4</b>

This means a difference of almost 4 grade levels by the end of middle school.

achievement. Finally, if your child attended a school averaging 140 percent, he would graduate at 10.4 years of academic achievement — that’s a difference of almost four full years of achievement over a period of seven years.

The policy implications that arise from an understanding of value-added are highly significant once one grasps the critical distinction between the meaning of student achievement and student growth. *Achievement* describes the absolute levels attained by students in their end-of-year tests. *Growth*, in contrast, describes the progress in test scores made over the school year. The best predictor of an absolute score such as a 1500 on the SAT is family income. But if we are predicting student growth — progress made over the year — *good instruction is 15-20 times more powerful than family background and income, race, gender, and other explanatory variables*. We are vastly better off, for example, investing in our educators through effective professional development than in reducing class size.

### Value-Added as a Diagnostic Tool

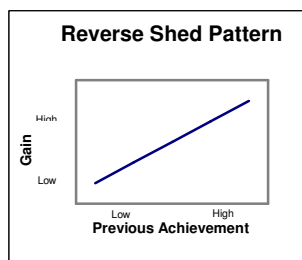
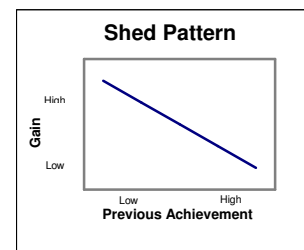
Value-added makes it possible to provide educators with data that allow them to determine the *focus* of their instruction (identifying which students have benefited most) and their instructional *impact* (how effective it has been in providing students with a year’s worth of growth from where they began the year).

Through this information, teachers, principals, district administrators, and school board leaders can learn whether previously high achievers, middle achievers, or low-achievers are making the most progress, and the extent to which schools and classroom teachers are effective in raising performance.

#### *Instructional Focus*

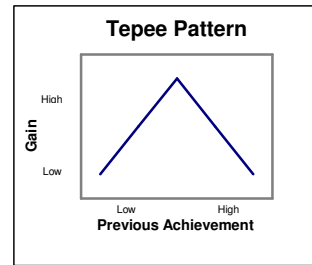
Using the extensive pool of data from Tennessee collected by his EVAAS® system, Sanders has identified three distinct classroom patterns, whose names reflect the shape of the slopes in the descriptive charts. These patterns can be found in any classroom, school or district, but occur disproportionately in the circumstances described below.

The *shed* pattern explains the frequent observation made by teachers in low-income communities: the incidence of precocious students falls precipitously from first through sixth grade. Teachers in these schools, faced with so many low performing children, focus on the bottom of the student distribution so that previous low-achievers get high gain while previous high-achievers get low gain. Sustaining this focus in the early years results in few high achieving children by middle school.



The *reverse shed* pattern reveals the opposite focus that is found disproportionately in high-income communities. Here teachers appear to respond to the central concern of their communities by concentrating on their highest performers. As a result, previous low-achievers get low gain while previous high-achievers get high gain.

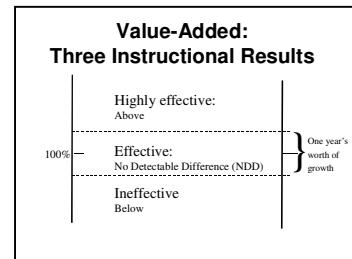
The final pattern is the *tepee*. In these classrooms, teachers are concerned with the average performer. By focusing on the middle of the distribution, both previously low- and high-achievers get low gain, while the previously average-achievers get high gain.



*Instructional Impact*

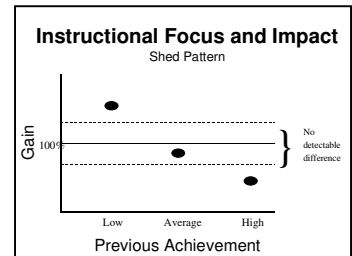
To describe the impact of instruction on student learning, value-added assessment calculates three-year running averages for the value-added gains made by all students in individual classrooms. These data can then be aggregated so that impact can be determined as well for grades, schools and districts.

We refer to these instructional results as: *Highly effective*: students are “stretched” so that their performance significantly exceeded their records of past achievement; *Effective*: students received a year’s worth of growth from where they began the year; and *Ineffective*: student performance consistently and significantly fell below the level of achievement they demonstrated in past years.

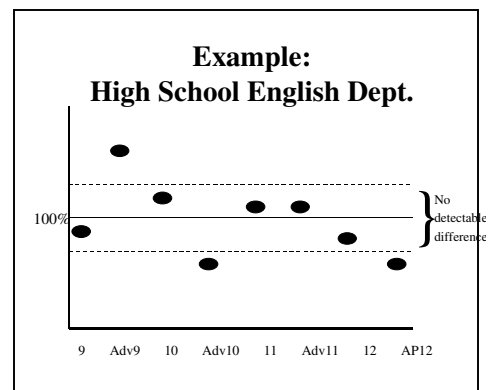


When the data from these patterns are combined, educators can see simultaneously the focus and impact of their instruction.

In evaluating the combined patterns, educators must always keep in mind the achievement levels before deciding whether the value-added patterns are appropriate or whether they need to be changed. Teaching in a shed pattern, for example, means one thing if the instruction is effective for all groups of students and quite another if the instruction for all groups is ineffective.



In the particularly interesting illustration below, the chairman of a high school English department examined the value-added performance of eight teachers in grades nine through twelve who taught either a regular English class or an advanced class. In five of the eight classrooms, students were receiving a year’s worth of growth in a year. But, in the ninth grade advanced classroom, previously high-achieving students were being “stretched” academically and in the tenth grade advanced classroom, they were losing ground. Since the teachers in the advanced ninth and tenth grade classrooms were assigned because the chairman believed them both to be excellent instructors, he found the results puzzling. I use this example as a caution. The easy explanation is that there is something wrong with the tenth grade teacher’s classroom instruction. That turns out, however, not to be the case. The real problem was that the tenth grade instructor didn’t communicate with his colleague teaching ninth grade and hence didn’t know that these students were performing at exceptionally high levels. As a result, he spent a good part of the fall term



reviewing what these students had already mastered, thus depriving them of the growth to which they were entitled. This means that good instruction must also extend to behaviors outside rather than simply inside classrooms.

The diagnostic value of these patterns of instructional focus and instructional impact is enormous. Imagine a session at which all six of a school's fifth-grade teachers convene to discuss the efficacy of their math instruction. Everyone present has a copy of each other's value-added patterns and whoever serves as the instructional leader begins by asking different teachers — the highly effective, those most successful with previously low-achieving and high achieving students — to explain their pedagogical strategies. The conversation continues when someone suggests an article for everyone to read for next week that sheds light on the precise issues with which they're struggling. The chances that math instruction will improve in that school have just been raised significantly.

Not differentiating instruction, limiting problem-solving opportunities in the curriculum, or failing to communicate with colleagues who taught their students in previous years — all of these practices deprive students of gains they are capable of making and to which they are entitled.

The point here is that value-added assessment by itself does not improve student achievement. But if educators are committed to analyzing the valuable data it provides and use what they learn to guide instruction and professional development, and if administrators create an environment that encourages these activities, more students will be able to achieve at higher levels.

So now let's summarize several features of value-added that catalyze high-quality instruction. First, it provides unique data about each teacher's classroom. It is perhaps the most "job-embedded" expression of professional development — that's the jargon that appears repeatedly in the literature arguing that to be useful training must be directly relevant to the classroom teacher.

Second, it ends the isolation of teachers and teaching. In one of the best articles written about K12 reform — "Building a New Structure for School Leadership" (Albert Shanker Institute: Winter, 2000), Harvard's Richard Elmore describes public education as consisting of two "loosely coupled" parts: a "core" (teachers and students) and an "administrative superstructure" (school boards, superintendents, principals, etc.). The culture of the core, according to Elmore, socializes teachers to believe that when they close the door to their classroom, what goes on between them and their students is personal and intimate, and if you are given to a romantic interpretation, even a bit magical and mysterious — but it is fundamentally private. This is good and as it should be. Indeed, it is the responsibility of the administrative superstructure to buffer the core from outside interference. The unintended consequence of this culture is the isolation of the teacher, and isolation, as Elmore explains, is the enemy of learning. When teachers convene regularly to talk about the patterns of their instruction, isolation ends.

Third, "learning communities" form as teachers begin to read about ways to change their patterns in desired directions and to meet regularly with their colleagues to discuss what they have learned. In short order, we expect professional development vendors — for profit

companies as well as area colleges or intermediate units — to develop subject- and grade-specific modules to facilitate these ongoing conversations. These modules will contain the appropriate questions and readings and will sequence them to sustain these inquiries over a school year.

Fourth, school leaders as well as teachers will have the requisite data to drive decisions about curricula and professional development. We are finally at the moment in education history when researchers have the right data and technology to undertake analyses that will yield critically valuable results about what does and does not work in classrooms. Studies can be constructed that hold constant the achievement levels of students and the value-added effectiveness of teachers to learn which among competing curricula modules or professional development interventions are most effective.

Finally, value-added will have a highly positive impact on the morale of teachers working in low-income, high minority, or working class school districts if they do a good job in providing their students with a year's worth of growth or more. In other words, districts that historically have ranked low in standings based on achievement tests can compete for the first time on a level play field with schools in wealthier communities because their rank will be determined by a value-added score rather than by absolute scores that are highly influenced by family income.

A few more important issues need to be addressed before closing. Before a state or district can implement value-added assessment the following things need to be in place. You need unique individual identifiers for students and teachers, and you need assessments that are closely aligned with the standards and curricula. You also need annual tests that are available in “fresh, non-redundant and equivalent forms.” Each of these terms has a technical meaning, but overall this guarantees that when these conditions are met, the test cannot be “gamed.” That is to say, no advantage would be given to a teacher's students if they practiced using last year's test. It is also preferable, although not absolutely necessary, that the tests be vertically scaled.

A second issue is the importance of collecting classroom data where value-added assessment is being adopted for statewide use. This is not the case currently in Pennsylvania and Ohio and it is most unfortunate. When the Pennsylvania State Board of Education mandated value-added for use in all school districts, it was highly impressed with the fact that the system provided a means of tracking individual students. This would make it easy to identify struggling students and to get them the help they needed to improve and thus help the state's schools meet their Adequate Yearly Progress goals. The Board was unaware that it was making an assumption that struggling students were found randomly in classrooms.

While struggling students are indeed found in classrooms of all types, the Tennessee data make unequivocally clear that they are not randomly distributed: they are found disproportionately in classrooms with ineffective instruction. If focus is on only the struggling students rather than on the teachers who are providing ineffective instruction, scarce resources will be devoted to the symptoms rather than their underlying causes.

This is well illustrated in a parable. Seeing a baby floating down a river in a basket, a man jumps in to pull the infant to safety. He spies another and yet another baby floating towards him. He cries for help bringing his neighbors, but soon several leave. “Where are you going?”

he calls, furious at their departure at a time of great need. “We’re going upstream,” they reply, “to find out who’s throwing the babies in!” Unless we get ineffective teachers the help they need to improve their instructional practices, we will always be dealing with symptoms not causes.

As the data on teacher effectiveness make clear, raising student achievement requires direct focus on the classroom where learning actually takes place, and value-added assessment provides us for the first time with a tool to render a fair and objective evaluation of classroom instruction. For this reason, value-added can also be the foundation for a system of accountability that holds individual teachers and administrators responsible for student-learning results. Let’s be clear – no educator should ever be evaluated solely on the basis of a single measure, not even one as powerful as value-added. But in order to achieve significant improvements in student achievement, educator evaluations should be linked at least in part to student-learning results.

Some argue that such new systems of accountability are not necessary because once educators are armed with value-added assessment, significant change will be forthcoming; however, the Tennessee experience casts some doubt on this promise. By Sanders’ reckoning, districts that embraced value-added saw measurable improvement in student growth, but many others ignored it.

This may have resulted from the fact that the State Board regulations for implementing value-added limited its use to no more than eight percent of a teacher's evaluation. Another factor may have been the failure by the State Department of Education in the five years following its introduction to provide districts with appropriate professional development to help educators use value-added.

It is also quite possible that the accountability provisions put in place by NCLB will catalyze far more responsiveness among educators. In an environment that requires schools to ensure that all students reach proficiency, more educators may embrace value-added assessment and make full use of the powerful diagnostic information it provides, thus obviating the need for additional accountability. Time will tell.

What is clear however, is that value-added assessment offers two crucially important benefits. It is a far more accurate way to measure the performance of schools than absolute test scores. And though not a panacea, value-added provides educators with an unprecedentedly rich opportunity to improve classroom instruction.

Whether one concludes that it is best to use value-added assessment as a diagnostic tool, or as the cornerstone of an accountability system — or both — let there be no mistake: it is the single most powerful tool available to educators for measuring student progress and the effectiveness of instruction and instructional programs. Without it, we will continue to work in the dark, tinkering at the edges of the system, and ignore what has always mattered most: good teaching.

**APPENDIX I**  
**OPERATION PUBLIC EDUCATION:**  
**ASSESSMENT AND ACCOUNTABILITY SYSTEMS TO TRANSFORM AMERICA'S SCHOOLS**

Operation Public Education (OPE), a foundation and corporate funded school reform effort based at the University of Pennsylvania, has developed a new and comprehensive accountability system that complements the federal law and provides tools to help schools meet their NCLB goals.

**Value-Added Assessment**

Value-added assessment serves as the foundation for the reforms. Since this methodology is discussed at length in the body of the paper, it is not described here other than to say that in isolating the impact of instruction on student learning, it provides an empirical basis for evaluating teacher effectiveness.

**Educator Evaluation and Compensation**

OPE's accountability system is based on a balance between teacher efforts and student learning. Teachers' in-class performance will be rated by their peers (rather than administrators) using the new protocols developed by the Association for Supervision and Curriculum Development that cover planning and preparation, classroom environment, instruction, and completion of professional responsibilities. To assess student learning, teachers will be evaluated on the progress their students make measured by value-added assessment. A teacher's value-added score will be calculated based on the progress made by all the students in their classes, averaged over a three-year period. The system handles high rates of student turnover by attributing student's scores in direct proportion to the time they are in a given teacher's classroom.

Based on an equal weighting between observations and value-added results, teachers will be classified as advanced, proficient or unsatisfactory. Utilizing Peer Assistance and Review teachers who are unsatisfactory in value-added and/or their observations will be required to undergo remediation and given the opportunity and support to improve in a process jointly administered by teachers and administrators.

Unlike the current system of teacher compensation – which pays salaries primarily for teachers' years of service and graduate credits – the OPE system links compensation to job performance through the use of a career ladder. Teachers move up the rungs based on their evaluations. At each stage of the career ladder, teachers receive higher pay and increased leadership opportunities. Although the OPE system provides for minimum increases as teachers climb the ladder's rungs, local school districts will still determine salaries through collective bargaining. Additional pay within each rung will continue to be based on years of experience. OPE's system includes a "grandfather" clause so that current teachers can choose to remain in the old compensation system (though they are still subject to the new evaluation process), but all new teachers must enter the new system.

OPE's accountability plan addresses the role of administrators as instructional leaders by evaluating them on how effectively they promote high standards for all students, use student-learning data to make decisions, and create an environment for their staff centered around student achievement and continuous professional development. Administrators will also be compensated based on a career ladder that recognizes their skills and accomplishments (50%), their success in meeting the Adequate Yearly Progress (AYP) goals under NCLB (25%), and the value-added scores of their school or district (25%). They begin as *Interns* with a mentor administrator, then progress to *Career* stage and, if they demonstrate excellence, reach *Distinguished* status.

### **Professional Development**

Professional development – “in service” training given to teachers – is the indispensable link between educator accountability and student performance. For students to meet demanding standards, schools and districts need to provide their educators with access to the skills that will help them become better teachers and administrators.

OPE’s model ensures that educators are engaging in ongoing training that is linked to the learning needs of students. The system calls for appropriations for 12 days of professional development per year and the creation of new categories of teachers recruited from the ranks of Advanced and Distinguished educators: *mentors* to work with new teachers for several years and *coaches* to work with colleagues to improve their craft, resourced on the basis of one coach for every 300 students.

### **Capacity-Building**

Success requires access to adequate resources and ongoing professional training. OPE’s system recognizes the unique role played by regional units (such as Pennsylvania’s IU, New York’s BOCES or Ohio’s ESC) and strengthens them to provide needed assistance to school districts in their areas.

As required under NCLB, struggling schools and districts — those not meeting their AYP — will be assisted by “technical assistance teams” to help districts implement their school improvement plans. The OPE system strengthens these teams by ensuring that they are staffed by distinguished administrators and teachers. And instead of the “perverse incentives” that penalize or reward schools that are failing by withdrawing or adding funds, the OPE system gives technical assistance teams the authority to administer new funds and set policy directions to ensure that struggling schools get the help they need to improve.

### **A New and Expanded Role For Teachers**

OPE’s system creates a new pathway to deliver what most teachers have long wanted – improved school leadership, better working conditions, more valid evaluation, meaningful professional development, higher student achievement, and the opportunity to earn higher salaries.

The system improves the position of teachers by creating a new “quid pro quo” where teachers accept accountability as individuals in return for a significant expansion of their role in public education. First, since the classroom rather than the negotiating table will determine career advancement, teachers get an equal say in all major issues that affect classrooms such as curricula and professional development. Second, teachers rather than administrators are given responsibility for the observation portion of a teacher’s evaluation. Finally, when either through observation or value-added scores, a teacher is flagged for mandatory remediation, teachers are given four votes on a seven-member panel where a five-vote majority is required for all decisions. This ensures that all teachers are treated fairly while at the same time not allowing genuinely poor teachers to remain in the classroom.

### **A New System of Integrated Assessments**

The final element in OPE system is now being developed. It is a “Request for Proposals” that would be issued by states and large school districts. The RFP would include the criteria for an integrated assessment system that would provide not only a high-quality “summative” exam for accountability purposes at year’s end focused on the development of higher-order thinking skills, but “formative” assessments throughout the school year designed to give teachers regular feedback in the form of suggested pedagogical interventions to support improved instruction for this year’s students. Publishers have been reluctant to design such an assessment system until they can be assured that a substantial market exists for this product. OPE will work with key education stakeholder groups to build this market at the state level by promoting a high quality Request for Proposals (RFP) that states can then issue.