

**How value-added
assessment can be used
to drive learning gains**

MEASURING

WHAT MATTERS

By Ted Hershberg, Virginia Adams Simon, and Barbara Lea-Kruger

When the history books are written, “value-added assessment” will be understood as a revolutionary breakthrough—one that gave educators a powerful diagnostic tool for measuring the effect of pedagogy, curricula, and professional development on academic achievement and gave K-12 education a fair and accurate foundation on which to build a new system of accountability.

Value-added assessment comes at a critically important time. In a radical departure from past practice, the federal government is now requiring states to test every child across most grades and key subjects. No Child Left Behind is a welcome catalyst for fundamental change, but its potential will not be realized because of serious design flaws. One, in particular, is its use of raw test scores as a blunt instrument to determine progress for every school and school subgroup.

However well-intentioned, this approach has severe limitations and is unlikely to help school districts transform teaching and learning in every classroom. Raw achievement scores reveal in bold strokes how well broad groups of students are performing but do not provide information that can help determine whether students are making progress as a result of what the school or teacher is doing.

These assessments don’t follow students over time, don’t measure the independent contributions made by students and educators in the learning equation, and mask what is really happening in schools and classrooms by failing to eliminate key nonschool factors, such as family background or socio-

economic status, on student performance.

The data provided by most current accountability systems give us only the broadest view. If we are going to move all students to high standards and win public support for added investments in education, we need accountability systems that provide comprehensive, detailed, and accurate information that goes beyond raw numbers and can help transform teaching and learning.

Thanks to researchers such as William Sanders, formerly a statistician at the University of Tennessee, such an effective approach to measurement and accountability now exists and is being quietly implemented in states and communities nationwide.¹ Value-added assessment helps school decision makers determine how effective teachers and schools are, how to differentiate truly exceptional changes from predictable ones, and how to use data at the classroom level to make necessary adjustments in pedagogy, curricula, and professional development to bolster learning gains for every child.

This pioneering approach has been used across Tennessee since 1992 and was recently adopted for use statewide in Pennsylvania and Ohio. It is now being used in more than 300 districts and consortia nationwide.

Sorting out the factors of success

What is a good school? Educators would offer a variety of answers, but in the eyes of the media, the public, and many elected officials, a good school is defined by high test scores. The

problem is, test scores by themselves tell you a lot more about the students than about their schools because the scores are so strongly influenced by family income. The real question, then, is: Do good schools make good students, or do good students make good schools?

Value-added assessment provides the necessary tools to address that question, disentangling the independent contributions that both students and teachers make to student achievement. This enables schools to gauge the quality of instruction and eliminate extraneous factors (such as socioeconomic status) for low performance.

The value-added approach to assessment—and Sanders' philosophy of education—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 level—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 Sanders' value-added methodology.

A sound statistical approach

“Value-added” is often confused with simple growth scores—how much progress a child has made from one year to the next. But the statistical method known as value-added assessment tells us much more. 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.

Under the value-added approach, test scores are projected for students and then compared with the scores they actually achieve at the end of the school year. Classroom scores that exceed projected values suggest that a teacher is highly effective. Conversely, scores that are mostly below projections suggest that the instruction was ineffective.

But at the same time this approach considers student-related 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 previous academic history in the same classes have done well, this is evidence of the student effect—external variables such as the home environment that can be ruled beyond the range of a teacher's influence.

Value-added assessment is based on a sound—but statistically and computationally complex—methodology.² There is no simple way to isolate the impact of teaching on student learning. The complexity of value-added measurement (which requires mixed models and covariant structures) should hardly be considered a drawback, however. Not everyone who uses a personal computer or television remote needs to understand the intricacies of microengineering; similarly, not everyone who uses value-added assessment needs to understand its

complex methodology.

Numerous studies examining legitimate theoretical concerns about this methodology have found those concerns to be statistically insignificant. Moreover, extensive field testing of the Tennessee model that has identified and quantified these underlying classroom and school effects has continually demonstrated the assessment model's accuracy and reliability.

It is important to note that the value-added approach has proved to be a fair and accurate measure for students and educators alike. By using data from multiple years in multiple subjects, it considers the whole child, providing a more accurate analysis of student achievement. In addition, because students' projected scores are based solely on their prior academic record, value-added does not explicitly consider students' race or socioeconomic background, and no bias is introduced. In other words, the stereotypical expectations that low-income minority children will do poorly and wealthy white students will do well have no place in this system.

The system is fair to educators as well, precisely because, in basing a projected score on prior academic achievement, value-added assessment takes into account the bias inherent in absolute test scores. It is also fair for educators in schools with high student turnover because scores showing the teacher's effect on learning can be weighted by the percentage of time the student was in a particular teacher's classroom. In turn, accounting for students becomes a problem only when they leave a district, which is why this system works better in large school districts and best at the statewide level.

A powerful diagnostic tool

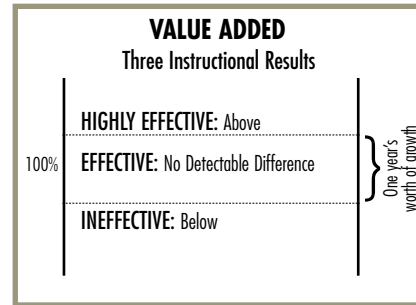
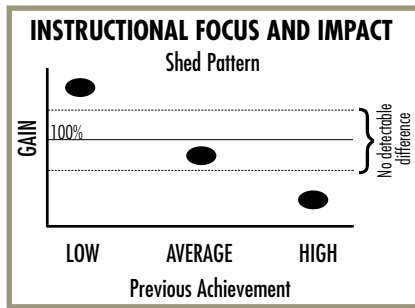
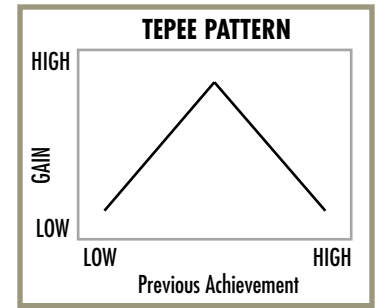
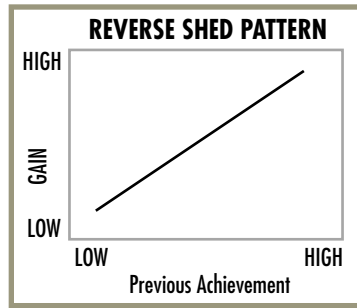
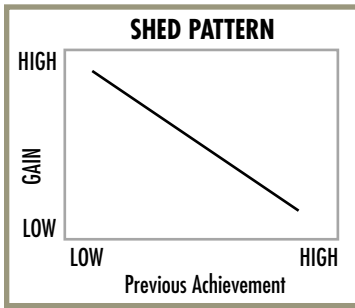
Studies have documented the powerful relationship between the quality of classroom instruction and student achievement. Family income remains the strongest predictor of scoring 1500 on the SAT—an absolute test score. But when predicting student progress, teacher effectiveness is 10 to 20 times more powerful than income, class size, race, or family educational background.³

Sanders, now manager of value-added assessment and research for the North Carolina-based SASinSchool, has used his research to develop diagnostic tools that allow educators to determine the focus of their instruction (identifying which students have benefited most) and their instructional impact (how effective they have 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 high achievers, middle achievers, or low achievers are making the most progress and how effective schools and classroom teachers are in raising performance.

Instructional focus

Using the extensive pool of data from Tennessee, Sanders has identified three distinct classroom patterns, whose names reflect the shape of the slopes in descriptive charts. These patterns can be found in any classroom, school, or district but



occur disproportionately in certain circumstances.

The shed pattern explains a frequent observation 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 students at the bottom, resulting in high gains for previous low achievers and low gains for previous high achievers. Maintaining this focus in the early years results in few high-achieving children by middle school.

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

The final pattern is the tepee. In these classrooms, teachers are concerned with the average performer. By focusing on the middle of the distribution, previously low and high achievers alike show low gains, while the previously average achievers show high gains.

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 impact can be determined for grades, schools, and districts. These instructional results fall into three categories:

- **Highly effective.** Students are “stretched” so that their performance significantly exceeds their records of past achievement.

- **Effective.** Students attain a year’s worth of growth from where they began the year.

- **Ineffective.** Student performance consistently and significantly falls below the level of achievement students demonstrated in past years.

When the data from these patterns are combined, educators

can see the impact of their instruction and how they may be focusing on some students to the detriment of others. Not differentiating instruction, limiting problem-solving opportunities in the curriculum, or failing to communicate with colleagues who taught their students in previous years, thus spending too much time reviewing what students have already learned—all of these practices deprive students of gains they are capable of making and to which they are entitled.

By seeing where and why they are effective, teachers can also share best practices with their colleagues. Good teachers may be highly effective with some students and quite ineffective with others. Understanding areas of strength can help correct areas of weakness.

In addition the data can have a significant impact on ways districts can align their effective and highly effective teachers to maximize student achievement. (For an example, see the sidebar on page 30.)

Value-added assessment by itself does not improve student achievement. But if educators analyze 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.

Achieving the goals of NCLB

NCLB requires schools to help each student reach prescribed proficiency levels. Schools without value-added data collected at the classroom level operate on the unspoken assumption that struggling students are randomly distributed in the population. But the Tennessee data reported in the sidebar make it absolutely clear that struggling students are found disproportionately in classrooms characterized by ineffective instruction. Helping students raise test scores without bolstering the quality of instruction is addressing the symptoms without dealing with the underlying cause.

The reality in roughly half the nation’s schools is that a significant gap exists between where many students are now and

where NCLB says they need to be. The gap must be closed each year, and measurement must be done using absolute test scores—you either hit the target or fail.

Value-added enjoys several advantages over Adequate Yearly Progress (AYP) measures. For one thing, it traces individual students over time, while AYPs use cohort comparisons—this year's fourth grade is compared with last year's fourth grade. Cohorts do not allow you to see if schools are having an impact on the same students over time. A school could be on the warning list one year and successful the next, largely because a different group of children is being tested.

Second, because it measures student growth, value-added can distinguish schools that are making no progress from those that may have missed their absolute targets but are clearly making some progress.

Finally, value-added maintains a focus on *all* children. Under NCLB, the inevitable focus of schools is on the subset of students whose improvement will satisfy their AYP goals. Thus students performing above proficiency and those far below it are more likely to be ignored in the process.

But the law is the law, and if schools are to close the proficiency gap over time, they must do two things. First, they must not lose value. Tennessee data make clear that far too many schools actually lose value annually. In 1996-97, for example, a large majority of the state's schools did not provide first-through eighth-graders with the year's worth of growth in math to which they were entitled.

Second, schools must make the necessary investment to build capacity and provide strong instructional leadership. This means recognizing the unique catalytic role that value-added assessment plays in enabling "job-embedded" professional development, ending the isolation of teachers and teaching, building learning communities where educators are actively engaged in discussing what is happening in their classrooms, and enriching the process of data-driven decision making. This requires focusing every available resource on activities that support high-quality instruction.

Some critics of value-added assessment believe the focus on growth rather than on absolute test scores works to the disadvantage of low-performing children. Even though they may get a year's worth of growth annually, the argument is that these students are still likely to leave school below proficiency because they started at such low achievement levels.

That is why our answer to the question "What is a good school?" must satisfy two conditions. First, in any given year, a good school must provide at least a year's worth of growth on average for every student. Second, all students in all subgroups and grades must be stretched beyond their year's worth of growth to achieve proficiency as defined by a state's NCLB standards by the time they graduate.

The cornerstone of comprehensive accountability

No educator or student should ever be evaluated solely on the basis of a single measure, not even one as powerful as value-

THE POWER OF GOOD TEACHING: A TENNESSEE CASE STUDY

Since 1992, William Sanders' Educational Value Added Assessment Services (EVAAS) has recorded each Tennessee student's scores on TerraNova exams in every subject and in every grade linked to the teachers who taught them. From these data, value-added scores (measures of instructional effectiveness) have been calculated for individual teachers.

June Rivers, associate director of EVAAS, ranked all fourth-graders in two large metropolitan districts in Tennessee and divided 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. The results reported here are for the bottom quartile, but she found the same patterns in each quartile.

Because 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 that student passing the test were less than 15 percent.

Explanations for this result typically rely heavily on factors beyond the school. Some observers cite difficult socioeconomic

circumstances. Others point to 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," we are told. "You can bring the horse to water, but you can't make it drink." These explanations share a deeply flawed conclusion: that the fault lies with the students rather than with the quality of instruction they received.

Yet, if these same students had 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 truly exceptional top 1 to 5 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 who are considered hopelessly low achievers or academic failures can perform at much higher levels—if they receive high-quality instruction.—*T.H., V.A.S., and B.L.-K.*

Source: June Rivers, "The Impact of Teacher Effect on Math Competency Achievement," unpublished doctoral dissertation, University of Tennessee, Knoxville, 1999.

added. But to achieve significant results in student achievement, educator evaluations should be linked at least in part to student learning. 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 a tool to render a fair and objective evaluation of classroom instruction.

For this reason, value-added can also function as the cornerstone of an accountability system that holds individual teachers and administrators responsible for student-learning results.

Some argue that such a system is not necessary: Once educators are armed with value-added assessment, they say, 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 might be because state regulations for implementing value-added limited its use to no more than 8 percent of a teacher's evaluation. Another factor might have been the failure by the State Department of Education to provide districts with appropriate professional development to help educators use value-added assessment.

It is also quite possible that, under NCLB's requirement 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.

In any case, if schools are to graduate all students with the requisite skills for success in the knowledge-based society of the 21st century—the ability to use technology, solve problems, think critically, and learn on their own throughout their lives—the nation must be prepared to invest substantial new funds to groom top-flight instructional leaders and help them master the necessary new skills, pedagogy, and concepts. Think of it as a simple quid pro quo: carefully targeted investment in return for fundamental education reforms.

Operation Public Education, a multiyear effort funded by foundations and corporations and based at the University of Pennsylvania, has developed a new and comprehensive accountability system. Many of its elements are currently in place in states and school districts across the nation. The great strength of the OPE approach is its unprecedented comprehensiveness—the whole is indeed greater than the sum of its parts. Highlights of the system are described in four broad areas—assessment, capacity, professional development, and evaluation and compensation—and may be accessed at the Web site: www.operationpubliced.org.

Success in implementing this new system of accountability will require a second quid pro quo, one that school boards should embrace. In the new system, unions would be able to effectively represent the interests of their members only if they pay as much attention to the issues of quality of instruction and student learning as they have to salaries, benefits, and working conditions. If the unions accept a system of evaluation and

compensation that holds members accountable for their success in a classroom, school boards and administrators must share decision-making authority with them on all issues pertaining to professional development and curriculum.

In the NCLB era, 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, it gives educators 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, ignoring what has always mattered most: good teaching.

Ted Hershberg is a professor of public policy and history and director of the Center for Greater Philadelphia at the University of Pennsylvania, where Operation Public Education is based.

Virginia Adams Simon is the center's associate director, and Barbara Lea-Kruger is director of communications and development.

NOTES

1. Harold Doran's important article on value-added assessment, "Adding Value to Accountability," *Educational Leadership* (November, 2003, vol. 61, no. 3), appeared too late to be discussed in this essay. New American Schools, where Doran is director of research and evaluation, has developed its own method of value-added assessment, and there are others as well. Our essay focuses on the Sanders model, the only form of value-added assessment currently used on a statewide basis.

2. In addition to references cited in the Doran essay, see also David A. Harville, "A Review of Value-added Assessment System (TVAAS)," Iowa State University, 1996; Walter W. Stroup, "Assessment of the Statistical Methodology Used in the Tennessee Value-added Assessment System (TVAAS)," University of Nebraska, Lincoln, 1996.

3. See John F. Kain, "The Impact of Individual Teachers and Peers on Individual Student Achievement," *Association for Public Policy Analysis and Management*, Nov. 17, 1998; Eric A. Hanushek, John F. Kain, and Steven G. Rivkin, "Teachers, Schools, and Academic Achievement," *National Bureau of Economic Research*, 1998; and William Sanders and June Rivers, "The Cumulative and Residual Effects of Teachers on Future Student Academic Achievement," *University of Tennessee Value-added Research and Assessment Center*, 1996.