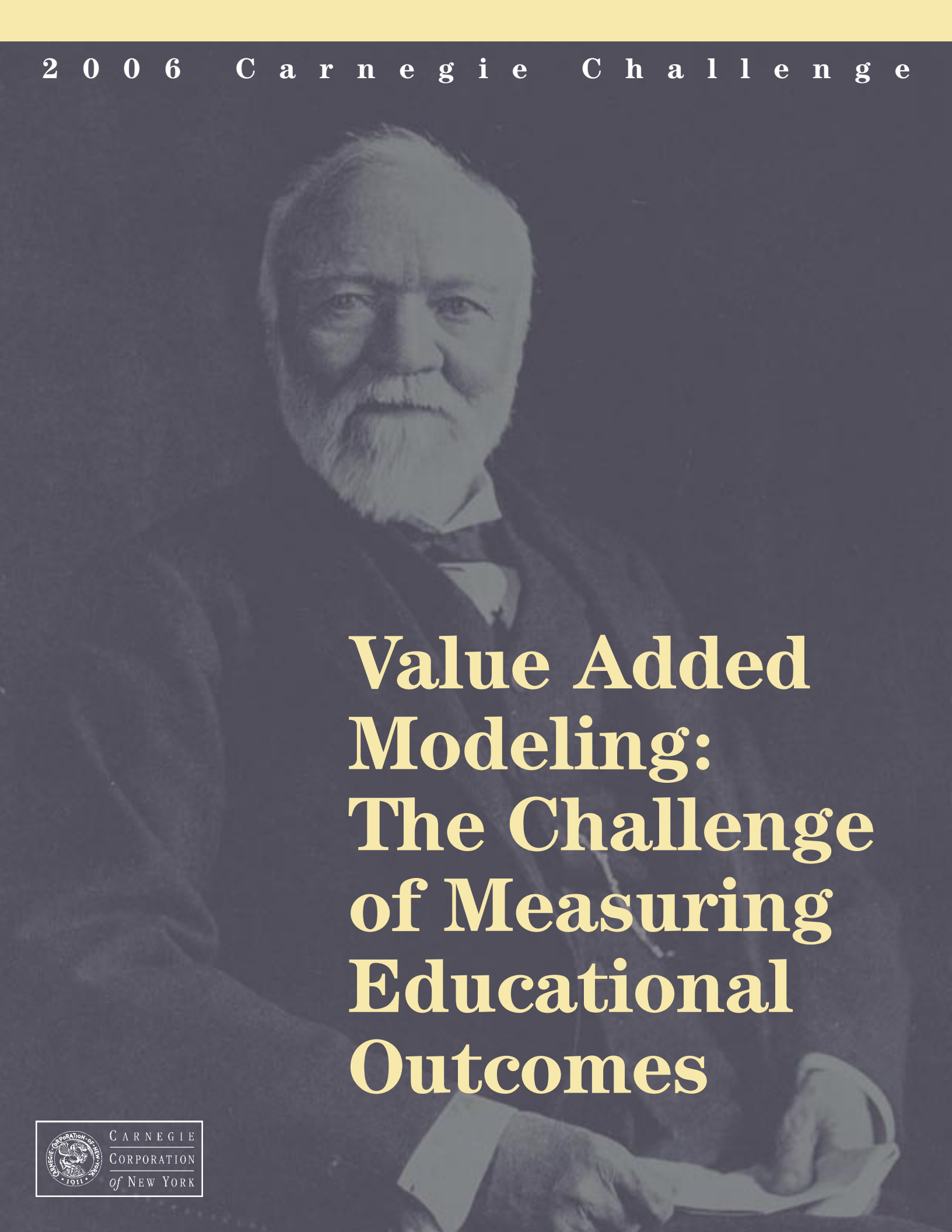


2 0 0 6 C a r n e g i e C h a l l e n g e



**Value Added
Modeling:
The Challenge
of Measuring
Educational
Outcomes**

Value-Added Modeling: The Challenge of Measuring Educational Outcomes

by Barbara Elizabeth Stewart

ABOUT THE AUTHOR

Barbara Elizabeth Stewart, a graduate of St. John's College in Santa Fe, New Mexico, and of Columbia University, has been a journalist for some twenty years. She has written for *The New York Times*, *The Washington Post* and the *Observer* and the *Telegraph* in Britain, among other publications. She lives in New York City and has a particular interest in education.

The goal of Carnegie "Challenge" papers is to lift up ideas and issues in a way that we hope will elevate them to the national agenda. The subjects we deal with, along with questions we explore, grow out of the work of Carnegie Corporation of New York but do not necessarily represent the focus of our programs. For more information about the Corporation's grantmaking activities, please visit our web site: www.carnegie.org.

Carnegie Corporation of New York was created by Andrew Carnegie in 1911 to promote "the advancement and diffusion of knowledge and understanding." Under Carnegie's will, grants must benefit the people of the United States, although up to 7.4 percent of the funds may be used for the same purpose in countries that are or have been members of the British Commonwealth, with a current emphasis on sub-Saharan Africa. As a grantmaking foundation, the Corporation seeks to carry out Carnegie's vision of philanthropy, which he said should aim "to do real and permanent good in this world." Currently, the foundation focuses its work in four program areas: Education, International Peace and Security, International Development and Strengthening U.S. Democracy.

Carnegie Corporation of New York • 437 Madison Avenue • New York City, NY 10022

What difference can a teacher really make? Is that difference measurable? Can teachers pull floundering students up, inspire them and enable them to achieve scholastically? Conversely, do they have the power to leave them flat, to frustrate or bore them into academic failure?

And further, can the skills of the best teachers be isolated, studied and transmitted to neophytes? If so, would that be a way to increase student achievement across the country, in schools located in both prosperous and poorer areas?

These are contentious questions and, until fairly recently, most educators would have answered by saying no, not really to all of them. Teachers, it was thought, did not matter much. What mattered most in student achievement were factors beyond teachers' control: primarily, the child's race, class and family income.

Certainly, there are always tales of remarkable teachers in some of the country's worst schools who manage to pull disadvantaged students to academic heights—to a mastery of advanced courses and, perhaps, admissions to competitive colleges. But by and large, those stories have been seen as inspirational exceptions—tales for *People* magazine, not education journals. They were considered irrelevant to the vast majority of teachers and schools.

But in 1982, a Tennessee statistics professor named Bill Sanders, experimenting with a statistical technique once used to breed sheep, began to change those ideas—not just in academia but in the world of schools, teachers and students. The method, an economic technique called value-added modeling, uses a complicated statistical system to compute data. It follows the academic achievements of individual students over several years and, according to its proponents, is not skewed by such factors as race, income or English

proficiency. But if the methods are complex, the concept is easily definable. “The basic idea is that you can use statistical models to see how good teachers are,” says Daniel McCaffrey, an education policy expert at the Rand Corporation.

Value-added modeling is a method of measuring student academic progress over time. It uses the annual standardized test scores for individual children, administered at the beginning and the end of the school year, to plot their progress in fundamental academic skills, and applies the results as a measure of the effectiveness of teachers and schools.

Lately, in an era of school accountability, value-added modeling has been attracting serious interest among education researchers and administrators.

“Parents often go into a closet and make height marks on the wall when their kid is growing up,” says Sanders, the former University of Tennessee statistics professor who developed and is now marketing his version of value-added modeling to many school districts. (Sanders has left Tennessee and is a senior research fellow with the University of North Carolina system as well as manager of value-added assessment and research for SAS Institute Inc. in North Carolina where he is assessing test scores sent to him from schools throughout the country.) “They plot the kid's growth. Well, we plotted the kids' growth patterns in math just like you would with their heights.”

Sanders was not the first to come up with the idea of measuring progress rather than straight achievement. Eric Hanushek, for instance, a senior fellow at the Hoover Institution of Stanford University, began working with value-added modeling methods in the early 1970s. But Sanders, who has devised his own statistical system, was the first to bring value-added modeling to a wider audience of educators and policymakers.

Value-added modeling works by attributing growth patterns higher or lower than average to teachers' skills—or the lack of them. When the method is used, many researchers, policymakers and educators are confident it provides strong evidence that teachers are the most important element in student achievement—no matter what the students' racial, class or economic backgrounds. A student who is assigned to a series of good teachers is more likely to achieve than one in a “better” school, in a more prosperous area, assigned several less effective teachers.

Though the degree to which teachers can influence student achievement is a matter of hot debate, it is almost universally accepted, now, that teachers can make a significant, and perhaps profound, difference.

Value-added modeling is one of the newest developments in the nationwide push for school accountability. Once as controversial as the importance of good teaching, accountability is now widely accepted as useful in increasing academic standards. It seems to be here to stay, at least for this generation. Since 2002, the No Child Left Behind Act (NCLB), of course, has mandated standardized testing in all schools receiving federal money and includes sanctions for schools that do not make the grade after a period of time.

But though NCLB has been promoted, in part, as a way of equalizing better-off, predominately white schools with low-income, largely minority ones, it provides no methods for improving schools that are lagging. Nor does it identify the teachers who are most effective—who deserve recognition and whose skills should be emulated. A major criticism of NCLB is that it mandates student achievement without offering methods for obtaining it.

To many education researchers and school administrators, value-added modeling may provide one solution. Some say that it could be the single most

promising way to identify effective teachers and effective schools.

“Over the next five-to-ten years, it could be the most important component of serious improvement strategies,” says Ross Wiener, who directs the policy team at the Education Trust in Washington, D.C., which focuses on education research and advocacy. “To get better outcomes,” Wiener says, schools “need to get better teachers and get teachers to be more effective.” Value-added modeling, he continues, lets administrators point out the problems and strengths of teachers and schools, and determine what supports they need.

More than a dozen states, including Florida, Ohio, New York, Pennsylvania, Colorado and California, are studying, and in some cases, applying, value-added modeling. Recently, Margaret Spellings, the Secretary of Education, announced that the U.S. Department of Education would accept applications from up to ten states to meet their part of their Adequate Yearly Progress (AYP), a statewide accountability system mandated by No Child Left Behind with value-added modeling. (The AYP requirements are currently based entirely on standardized test scores.) Beginning in the fall of 2006, the AYP in those states' schools will be calculated by using both the new progress method and the usual standardized tests.

The proponents of value-added modeling call its results fairer and more accurate than those produced by standardized tests measuring achievement only because socioeconomic factors may greatly affect the results of those tests. Value-added modeling can identify effective teachers and effective schools that are helping low-achieving children progress even when, as is common in low-income, minority districts, the students' standardized scores are below average and their schools are not meeting AYP requirements.

A few years ago, value-added modeling was viewed skeptically by African-American education leaders. Would it be used as a sop for low-performing schools? Would it be a way for district officials to duck the job of improving lackluster student achievement rates?

To the Congressional Black Caucus, “It was very important not to let schools and districts off the hook for producing a full measure of minority student achievement,” says Christopher Edley, dean of the University of California, Berkeley, law school, co-founder of the Harvard Civil Rights Project and a longtime education scholar. But now, he says, nearly all Congressional Black Caucus members approve of value-added modeling, as long as it doesn’t displace standardized testing requirements. In some states, African-American leaders have even approached school officials to ask that value-added modeling be instituted.

“The great promise of value-added assessment is that it rewards educators for improving achievement even if the improvement is not sufficient to reach a distant threshold of proficiency,” says Edley. “It rewards teachers taking the lowest performing students and moving them up, while the standards approach doesn’t give you any credit for that.”

On the other hand, value-added assessments can also point out teachers and schools in well-off, high-achieving schools that are coasting and not helping adequate students go farther. Most importantly, according to its proponents, value-added modeling can identify teachers and schools that help students move ahead academically, whatever their starting point. (The value being added refers to the amount learned by a student over the year, and thus, presumably, the value being added by any given teacher.) Studies done by Sanders and by other researchers indicate that teachers are more than just another important factor: they are crucial. Their influence on student achieve-

ment, positive and negative, can be enormous.

Some education researchers and academics are especially excited by the potential of value-added modeling to improve teacher education. Higher education officials could use the data to identify superior teachers, study them in the classroom and figure out what they are doing that is working so well. They could then include their techniques in teacher education curricula. In addition, value-added modeling could be used to make higher education institutions accountable for the effectiveness of their graduates in their first jobs.

Though research on value-added modeling in education began in the 1970s, it didn’t pick up steam until after 2000. It is the result of a convergence: of sophisticated computer technology, of the drive for accountability and of the work of education experts, including one—Bill Sanders—who was able to catch the interest of the Tennessee legislature early on. Tennessee, which began annual standardized testing decades before it was mandated by the federal government, provided years of data. New computer technology made it possible to process hundreds of thousands of test scores. The nationwide push for school accountability through testing provided the impetus. And now that all states are using standardized testing, the pool of data available for value-added modeling and other kinds of research is growing evermore vast.

What is value-added modeling currently being used for? Many school districts are planning to use it—and a few already are—to identify and reward effective teachers and effective schools. In some districts, the best teachers are acting as coaches, and the least effective ones are being given extra training. In addition, through the Teachers for a New Era initiative of Carnegie Corporation of New York, which is also supported by the Ford and Annenberg foundations, selected colleges and universities are using value-added

data to improve their schools of education and make them accountable for their graduates' performance in the field. They are starting to study the methods of the highest-scoring teachers so that they can be passed along to aspiring teachers. They are also beginning to follow their graduates into their first jobs, to find out how they are doing—and, if possible, to analyze why they are or are not succeeding—and to improve the work of those who need it. (See page 11.)

The result, if the hopes and plans are realized, will be a marked improvement in student achievement at a time when that seems to be badly needed.

History

Assessing the quality and achievements of teachers and of schools has long been a subjective and often haphazard business. Principals would talk to other teachers and tote up the number of complaints from parents. They made occasional classroom visits and noted the noisiest classrooms. “But were kids learning?” asks Daniel Fallon, chair of the Education Division of Carnegie Corporation. “Understanding *that* is what was missing.” The correlation between principals' ratings and children's learning was often remarkably low, Fallon notes.

Three important education studies of the late 1960s and 1970s concluded that teachers and schools could do relatively little to encourage student achievement. What mattered were race, class and income, all of which were well outside the control of teachers and schools. So definite were their conclusions that they became known as the “Schools Don't Matter” studies.

Equality of Educational Opportunity, widely known as “The Coleman Report,” a profoundly influential study commissioned by the federal government, was released in 1966 at the height of the civil rights movement. Some 800 pages long, it was intensely

thorough. Its primary author, James Coleman, considered to be the foremost mathematical sociologist of the day, used cross-sectional data comparing children of difference races, incomes and geographical areas. (It would be many years before computer technology was developed that would allow longitudinal studies of individual children.)

The Coleman Report, the first definitive study on the damage caused by racial and income inequities, became the impetus for the founding of Head Start and of continuing school integration. Its widely publicized conclusion read, in part: “Only a small part of [student achievement] is the result of school factors, in contrast to family background differences between communities.”

Inequality: A Reassessment of the Effect of Family and Schooling in America, released in 1972 by Harvard sociologist Christopher Jencks, and *How Effective is Schooling? A Critical Review of Research*, released in 1974 by the Rand Corporation, underlined Coleman's findings. The Rand study concluded, in part: “The most profitable line of attack on educational problems may not, after all, be through the schools.”

(Interestingly, the Coleman report briefly noted that achievement variations within schools were four times higher than those between schools. In many low-performing schools, in other words, significant numbers of students were doing very well while others in high-performing schools were doing poorly. Coleman speculated that the differences were probably caused by variations in teacher effectiveness, but Fallon notes that Coleman lacked the longitudinal data to pursue the point.)

It was 15 years before education reform was pushed to the top of the national agenda. *A Nation at Risk*, a report to the U.S. Secretary of Education by the National Commission on Excellence in Educa-

tion, released in 1983, was so strongly worded that it sent educators, parents and politicians into a near panic. It famously warned, “[T]he educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people,” and went on to say, “If an unfriendly power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen to ourselves.”

A Nation At Risk made education reform a burning issue. Several governors—Southerners, in particular—including Bill Clinton of Arkansas, Lamar Alexander of Tennessee, James B. Hunt, Jr. of North Carolina and Richard Riley of South Carolina, made education a priority. In the following years, education budgets grew significantly. Teacher salaries were increased, class sizes reduced and new facilities built. The drive for accountability began to grow, which, as a side benefit, provided a growing pool of test scores.

Around that time, Bill Sanders, a University of Tennessee statistics professor, was looking for a good, rich vein of longitudinal data—something that would make a meaty project for his graduate students. June Rivers, one of his students and now his professional partner, suggested that he try the standardized test data collected by the Hamilton County School Board, where she worked.

As it happened, Tennessee had started annual standardized testing in the 1970s, when an economic report warned that its low-performing school system was seriously hindering economic growth. Pressured by the business community—and some twenty-five years before NCLB made them federal policy—the Tennessee legislature mandated annual comprehensive academic testing. The tests, McGraw-Hill’s “Terra-

Nova” series, had an excellent reputation for high standards and reliability.

The accumulated scores provided Sanders and his students with excellent material for longitudinal research plotting the progress of thousands of individual students. The deeper Sanders went into the data, the clearer and more startling were his findings. He traced individual student scores over a period of years and compared them to one another and to the progress he determined an average student would make each year. He discovered there were wide variations among students of the same race and income. It seemed that students of some teachers—including a number in low-income schools with lower overall scores—were improving rapidly, while others in higher-scoring, prosperous schools were not.

The differences, he concluded, were attributable to the skill of the teachers. It seemed that the data were showing him that teachers could make a major difference in student progress, even among students who had been lagging far behind. The differences seemed to be independent of race and income.

Further value-added studies indicated that students assigned to three effective teachers in a row tended to make especially large strides, while those assigned to three ineffective teachers tended to fall far behind. (Sanders defined effective teachers as those with a majority of students who made greater-than-average progress over a period of time, and the ineffective ones as those whose students did worse than average over that period.)

Sanders concluded that his statistical calculations provided a way of measuring teacher effectiveness, in both high- and low-scoring schools and districts. It seemed that good teachers in good schools had the strongest influence on student achievement—welcome news, since the sprawling problems of race, class

and poverty were obviously beyond any one group's ability to fix.

Sanders spent a few years trying to explain his findings to local school officials, but eventually gave up. "I was totally convinced that nobody was interested in this stuff," he says. Then, one Sunday afternoon, he received a call from a newly elected young legislator, who had been given his phone number by a mutual colleague.

"He said he'd like to do something meaningful, quote, unquote, for education," Sanders explains. "My friend told him to call me." They talked for an hour-and-a-half. A few days later, Governor Ned Ray McWherter of Tennessee called Sanders and asked him to come on over to the governor's mansion and bring along those studies.

"They were looking for a different approach to accountability," Sanders says. "They bought into the notion that it's more fair to evaluate schools based on the progress rates of students than on standardized test results." That is how Tennessee began to use value-added modeling.

Over the years, Sanders has refined his methods. Basically, he devises an average progress level by pooling the test scores of students of the same grade, and then compares each student to the average. Using that information, he can plot the progress of students in a class, a school or a district.

As noted earlier, Sanders was not the first to come up with the idea of measuring student academic progress using longitudinal data. Hanushek, the Stanford University senior fellow, and Anthony Bryk, a professor at the Stanford University graduate school of business, were both pioneers in the area. But Sanders was the first to focus on longitudinal studies in-depth and persuade school officials to apply what was learned.

"Sanders got on his soapbox and sold this," says

Daniel McCaffrey of the Rand Corporation, who co-authored its value-added study. "He made a point of what he'd learned, that you can understand a lot about teachers from this kind of data. He got the information into the state legislature, and they got it to people who began really using it with teachers."

"The timing was right," McCaffrey continues. "It had large ripple effects. The significance of the data spurred others to replicate his study or do similar things. The notion that teachers and teacher accountability were important got people talking about value-added."

Nevertheless, Sanders has been criticized for his methods and his reluctance to publish them in the major academic journals, where they would be subject to peer review. "Bill hasn't done things to invite close scrutiny of his methods," says Dale Ballou, an associate professor of public policy and education at Vanderbilt University who, with Sanders, co-authored a report on value-added modeling. "He is secretive and private. Many scholars are unhappy with his methods."

Still, Ballou—and, it seems, the majority of education experts—regard value-added modeling as highly promising. By most accounts, it is the next important new development in education, one that many expect to become more solid and reliable as techniques are refined.

Discoveries

The findings of the experts in value-added modeling—including Sanders and researchers like Robert Mendro, chief evaluation officer at the Dallas Independent School District, who has devised different statistical methods—are unquestionably striking. And though some experts are critical of value-added modeling in general, and Sanders' methods in particular, as well as his preference for keeping his techniques rela-

tively secret, even the critics find considerable merit in the concept.

Essentially, proponents hold, value-added modeling underlines the impact of good teachers and of poor ones. According to the studies, having several good teachers in a row can greatly raise student achievement. The opposite is also true: one bad teacher can stall a child's academic growth and several can cause it to plummet.

According to *Teachers, Schools and Academic Achievement*, a 2002 report by Steven G. Rivkin, Eric A. Hanushek and John F. Kain, "Having high-quality teachers throughout elementary school can substantially offset or even eliminate the disadvantages of a low socioeconomic background."

In 1996, Sanders and Rivers reported that students of the most effective teachers scored an average of 50 points higher (on a 100 point test) than students of the least effective teachers.

Discouragingly, according to the Education Trust, low-achieving students were more likely to receive ineffective teachers and therefore tend to sink ever lower. But when they were assigned to a series of good teachers, their scores tended to rise substantially—so much that they often became above-average achievers.

In *The Real Value of Teachers*, published in 2004, the Education Trust reported that 100 percent of previously low-achieving Dallas students assigned to highly effective teachers for three consecutive years passed their comprehensive math tests. By contrast, fewer than half of the low-achievers assigned to less effective teachers for three years in a row passed the test.

Rand Corporation researchers, while upholding the basic thrust of value-added research, question the marked extent of the achievement differences between more and less effective teachers. In a detailed 2004 report commissioned by Carnegie Corporation and

entitled, *Evaluating Value-Added Models for Teacher Accountability*, co-authored by Daniel F. McCaffrey, J.R. Lockwood, Daniel Koretz and Laura S. Hamilton, Rand found a lesser degree of difference between effective and less effective teachers. Nevertheless, it said, "there is a growing consensus that teachers matter."

"Thus, the [Sanders and Rivers] results are consistent with the existence of persistent teacher effects but might overstate the size of such an effect," the report continues. Other similar studies, it added, "provide evidence that teacher effects do persist across the years."

The proponents of value-added modeling say the data are confirming that teachers matter enormously. But then, that's a fact that people with the most immediate, pressing interest in student achievement—parents and students—have always seemed to be aware of. There has always been jockeying and negotiations behind principals' doors to get the best teachers.

And just who are "the best teachers"? Most everybody at a particular school seems to have a pretty good idea.

There is, for instance, the ninth-grade English teacher, known for an ability to inspire and challenge students into emerging at the year's end with a solid understanding and appreciation for *Romeo and Juliet* and *Ethan Frome*. Year after year, this teacher's students are the ones with the most award-winning essays. Surprised parents might find their children reading Wharton with obvious interest and tossing out a Shakespeare quote at dinner. Even better, they might be delighted to learn that their previously lackadaisical scholars had pulled As and Bs in the class.

On the other hand, there is the teacher who turns Shakespeare's and Wharton's tragedies into dull, dry formulas to be memorized. Any parent who doesn't know this teacher's reputation has only to listen to the child's groans when the name comes up.

Picking out the very best teachers, the ones students remember with fondness long afterward and even credit with inspiring them into later accomplishments is rarely difficult. Choosing the worst ones, while perhaps more a matter of opinion, tends to be fairly easy, too. But what about the vast majority of teachers, the ones who range from the kind of bad to the very good?

The idea that teachers matter and can uniquely, strongly affect student achievement is becoming more widely accepted. According to its proponents, value-added modeling is the first reliable, objective method of separating the good teachers from the mediocre and poor ones. Without value-added data, they say, teacher assessment is subjective and often wrong. Judgments are based on personal likes and dislikes that lead to mistaken conclusions. Value-added modeling, supporters assert, provides an objective way of determining the effectiveness of teachers in performing their primary task: helping students learn and progress through the curriculum.

But even if the proponents are right, what should be done with the information? What is the best way to apply data that pinpoint the most effective teachers and schools?

Mendro at the Dallas Independent School District says that value-added data “is as effective as [how it is used by] the people who get the information. It’s sort of like measuring someone’s fever with a thermometer and then asking how effective the thermometer is.”

In Dallas, value-added data is shown only to teachers and superintendents. “In some cases, it’s used,” he said. “In others, it’s not. Mostly it’s not.” Making a similar point, Daniel Koretz, a Harvard education professor and one of the co-authors of the Rand report puts it this way: “Test scores describe. They don’t explain.”

How Value-Added Modeling is Being Used

How is value-added modeling being used? The broadest answer is: to measure and find the most effective teachers, schools, districts and curricula. About eight years ago, for instance, the Pennsylvania League of Urban Schools, which had been working to improve low-performing city schools, approached the state’s Department of Education with a proposal to institute value-added modeling. “Urban schools were being marked unsuccessful based on achievement,” says Gerald Zahorchak, the acting secretary of education for Pennsylvania. “The League felt their efforts were helping students make progress, so they wanted us to take a look at that.” They were especially concerned with the high percentage of urban students moving among schools, whose progress was not being followed. By measuring individual test scores each year, value-added modeling would track each student’s progress despite the moves.

For the past five years, Pennsylvania has been using Sanders’ value-added statistical techniques. The project, which began with 25 districts, has now expanded to 110. There are plans to institute it throughout the state.

At this early stage, officials are still learning to use the research, Zahorchak explains. Pennsylvania, at the beginning of a many-year process, is using the data to assess each district’s standards and ensure their alignment with state standards. Next, Zahorchak says, administrators will use value-added data to study the schools’ curricula and determine what is working and what is not. Only then will the state investigate the work of individual teachers and move into “intervention.”

“A lot of the time, it’s the teaching that’s bad, not the teachers,” Zahorchak offers. “Until you get the systems in place, it’s unfair to make judgments about in-

dividuals. Once the systems are aligned, an ineffective teacher will identify himself or herself. We should be able to get this done in a few years—then we can get close to addressing individual progress.” Eventually, he says, the salaries, promotions and contract renewals of Pennsylvania’s teachers and school administrators will be based partly on value-added assessments.

Another use of the method has been instituted in Pittsburgh. Faced with declining enrollment, officials have made school closure decisions based in part on value-added data. Elsewhere, states and districts farther along in the process are using it to pull up low-performing teachers and schools. Some of the measures used to help teachers with low scores are principal conferences, coaching and mentoring. Teachers with high scores may be appointed mentors or offered bonuses to shift to struggling schools. In addition, it is common to give bonuses to schools with high value-added scores, which are split among all the teachers, no matter what each individual’s scores are.

The most controversial uses are those that affect teachers’ livelihoods: raises, promotions, tenure and dismissals. Teachers’ union representatives on both the local and national level seem to be uniformly opposed to such uses. On the other hand, “A district can certainly use this information to pinpoint teachers for coaching,” says Maureen Peters, executive vice-president of the Alliance of Dallas Educators, a branch of the American Federation of Teachers. “Or if it’s a pattern, to [help them with] additional growth plans.”

A number of districts, including some in Denver, Dallas and throughout Tennessee, award bonuses to schools with high value-added scores. And, as noted above, these are divided equally among all the teachers, regardless of their individual scores. Peters says that is fine with the Denver union.

Officials in Tennessee, where value-added model-

ing has been used for nearly twenty years, believe that it has been instrumental in improving some of the state’s worst schools. A case in point is Hardy Elementary in Chattanooga/Hamilton County.

Some five years ago, it was one of the worst schools in the state, with standardized and value-added scores in the bottom twentieth percentile. Almost entirely African-American, it remains one of the poorest schools in a county that ranges from low-income to wealthy. Now, Hardy has the state’s highest value-added scores among elementary schools. Its value-added scores in math are the state’s highest, too. Its value-added scores in reading are in the top five percent and in other subjects, the top twenty percent. These notable improvements, it seems, are attributable to the combined efforts of the mayor’s office, the business community and the school district.

When Hardy’s low scores on both standardized and value-added tests were reported, the district appointed Natalie Elder, considered an especially dynamic administrator and former science teacher, as principal. The school was “reconstituted,” which meant that teachers had to reapply to work there. That gave Elder the chance to recruit some teachers of her choosing.

And when Hardy’s failing grades made news—in a county with some of the wealthiest neighborhoods in the state—Chattanooga’s municipal officials and business community rallied. They raised enough money to make it worthwhile for good teachers to work at Hardy, offering them \$5,000 bonuses and \$10,000 home improvement loans.

Kirk Kelly, director of accountability and testing at the Hamilton County School District, says Elder has focused on reading, math and teacher quality. “Expectations were very, very high,” recalls Linda Blazek, a comprehensive development teacher for

mentally handicapped students. “We were under a lot of pressure. It was hard for teachers. But we were the underdogs and wanted to tackle it.” As scores climbed though, so did morale. “Now we don’t think we can,” Blazek says. “We know we can.”

How central was value-added modeling to Hardy’s turnaround? Certainly, there have long been determined, effective principals who have raised achievement at troubled schools without the benefit of value-added data. And of course, if Chattanooga hadn’t acted, the value-added scores would have been just more proof of the obvious: that Hardy was in trouble.

Kelly says that value-added modeling was used to underline Hardy’s problems, to pick new teachers and to track its improvement. It provided feedback on what was working and what wasn’t. It also underlined the school’s success, raised morale among teachers and students and drew praise.

“When we see the results on paper, it makes us feel better,” says Blazek. “It’s positive proof for the public that we’re getting somewhere.”

Without value-added data, the praise would have been considerably muted. Hardy’s pass rate for the NCLB reading tests was 76 percent and for math, 69 percent. That is much improved, according to Kelly, though such scores would be unacceptable at high-performing schools in prosperous, white neighborhoods.

Lookout Mountain Elementary in Hamilton County, where houses cost between three hundred thousand and several million dollars, is one of those schools. Kelly cites it as an example of a school where the NCLB standardized scores are “wonderful” but the value-added ones are only “okay.” (However, according to some experts, value-added modeling does not assess the highest levels of achievement.)

It may take a long time for Hardy students to catch up to their advantaged peers at Lookout Moun-

tain, but Kelly and Blazek say that without value-added data, and the subsequent efforts of Chattanooga, Hardy would have remained at the bottom. Now they believe that the Hardy students have a chance.

Denver, which is newer to value-added modeling, is also planning on using the data for teacher incentives. After a four-year, \$9 million pilot program, value-added is being instituted citywide, says Brad Jupp, the senior academic policy advisor for the Denver public schools. Some sixty percent of Denver voters approved it in a referendum, despite the higher property taxes required for teacher performance raises. A similar percentage of teachers approved in a separate referendum, which meant that they agreed that one-third of their raises will be based on their students’ achievement. For low-scoring teachers, there will be coaching and other improvement measures.

“Routine teachers get routine pay,” says Jupp. “Exceptional teachers get exceptional money. Value-added modeling will make it easier to dismiss ineffective teachers who don’t improve.”

“The [field of education] has a notorious problem with dismissing teachers,” Jupp continues. “This information can lead to better decisions. Not that you need value-added data to dismiss a teacher. And it wouldn’t be the sole reason for dismissal. But it’s a very different decision if you have that information.”

Jupp credits Denver’s intensive study of the issue and its pilot program for winning the approval of the union and the majority of the teachers and the public. “We’ve had far less resistance than in most districts in the country,” he notes.

Some education researchers are especially excited by the potential of value-added modeling as a tool to rejuvenate teacher education at colleges and universities. Schools of education are “menacingly threatened,” says Carnegie Corporation’s Daniel Fallon. In-

creasingly, rather than wait for academically prepared teachers, school districts are seeking ways of getting teachers into classrooms quickly in programs that provide certification concurrent with full-time teaching, like the New York City Teaching Fellows.

As mentioned earlier, one notable undertaking using value-added data is Teachers for a New Era (TNE), a \$129 million Carnegie Corporation initiative involving 11 colleges and universities that has also received support from the Ford and Annenberg foundations. The TNE initiative aims to help colleges and universities extend their responsibility to graduates as they enter their first jobs, using value-added data to track and support their progress. Fallon says that one of TNE's most important design principles is the reliance on "persuasive evidence" of improved educational outcomes for students. What is essential, he says, is "making high-quality teaching visible through demonstrable student learning growth."

Critics

Advocates of value-added modeling are calling it the single most promising development in this era of academic accountability—perhaps the best hope for improving America's mediocre-to-poor public schools. But is that an overstatement? Is it at all accurate?

Every education expert surveyed for this report feels that value-added modeling is a promising development in recognizing good teachers and good schools. And none doubted the importance of good teaching. But as the Rand report made clear, what is not yet known is the magnitude of the teacher effects.

How much difference do good teachers really make? Is good teaching indeed the single most critical factor in student achievement, as value-added advocates say? Or is it simply one of many factors required by successful students?

Many critics question the advocates' conclusions. They do not necessarily believe that high value-added scores equal "good" teachers or that low ones identify "bad" teachers. In addition, they question the accuracy of value-added methods. They want much more research before applying value-added scores to teachers' salaries, promotions and contract renewals. They dislike what they consider overenthusiastic praise for it.

In other words, the critics of value-added modeling—who range from education professors and policy experts to union officials and teachers—want value-added's advocates to slow down.

"Not ready for prime time," is how Adam Urbansky, a Rochester teachers' union official, summed up his opinion of the value-added method, one that many experts have echoed. Some education experts, for example, point to what they say are flaws and weak points in the statistical methods of Sanders and others. These ardent advocates are overstating the advantages, critics assert. The critics, including the Rand report's authors, say that the percentage of error in value-added computation can be rather high, especially when compiled over three years or less and when the pool of scores is not large enough. (These errors apparently lessen when bigger pools of scores and longer periods are used in computations.) Therefore, say skeptics, it would be harmful to base teachers' livelihoods on what may be unreliable data. Doing so could result in low morale and in the end, be not much better than the old ways. (In fact, even the most enthusiastic advocates say that value-added assessment should be only one of several factors affecting teachers' raises and promotions.)

Sanders admits there is danger that necessarily complicated statistical methods will be over-simplified. Less expensive techniques may be substituted that will sacrifice accuracy. "It's a very sophisticated measure," explains Rob Weil, the deputy director of the educa-

tional issues department of the American Federation of Teachers. “It involves vertical scaling and sophisticated analysis to minimize errors. It’s not just: ‘Here’s what the kid got before; here’s what he’s getting now.’”

In Massachusetts, for example, growth measures are being used to compare the scores of one year’s fourth grade to the following year’s fourth grade. While that may provide useful information of some sort, it is not value-added modeling and should not be confused with it. It will not provide the same objective data on the quality of teachers, schools and districts. As Weil notes, “That doesn’t reflect how individuals do. It’s measuring this year’s apples with last year’s oranges.”

But as the idea of “growth” and “progress” measures become more popular, it is easy to imagine education officials, politicians and the public becoming confused about them and thinking that a simplified growth technique that costs less to apply would do just as well as proper value-added modeling.

The Rand report found much to praise in value-added modeling and encouraged research into it. But it had many reservations about the methods currently in use. For example, it criticized the standardized tests that some states are using and which are essential in devising accurate value-added scores. According to the report, the tests in some districts do not accord with state standards. Some tests being used do not assess the most- and least-advanced levels of knowledge, and some just do not provide accurate measures of student achievement. Further, the Rand report said that though the statistical techniques try to account for all the variables, they are new and not yet perfected and thus may produce significant errors. The report also notes that the factors affecting student achievement are complex and often have nothing to do with teachers’ effectiveness. As Daniel McCaffrey explains, they

may range from disruptive children or a wave of illness in school to the death of a principal.

Dale Ballou of Vanderbilt concurs. He thinks that value-added modeling in general—Sanders’ methods in particular, which he has studied—may inaccurately credit or blame teachers. As he explains, the value-added model dictates that if a particular student “does any better than you’d expect him to do based on his averages, you attribute that to the teacher. If he did worse, you also attribute it to the teacher, as well.” But other factors could cause a drop in a student’s achievement scores and that uncertainty, says Ballou, is value-added’s “Achilles heel.” He says, “This is the kind of thing teachers are worried about. What if they just get a class that’s going to be really tough to teach? There’s a lot of ‘luck of the draw’ in what kind of class a teacher is given.”

The likelihood of value-added assessment errors greatly increase with more complex curricula in higher grades, adds Daniel Koretz of Harvard. Furthermore, he points out, teachers who are effective in one classroom situation may not be in a different one.

While a number of noted experts agree that value-added modeling is most effective in pinpointing the most and least effective teachers, others suggest that “there’s a lot of noise in the middle.” Nevertheless, the Rand report agreed with the main conclusions of value-added research: that teachers do have an effect on student performance and that several good—or bad—teachers are likely to make a significant difference. The report concluded, however, that the difference cannot be accurately quantified using the data and methods currently available.

Koretz, one of the report’s co-authors, believes that value-added modeling can provide “valuable clues” and “vivid descriptions of what kids are learning.” But on the whole, he feels, it should be taken with the pro-

verbial grain of salt.

According to Stephen W. Raudenbush, professor of sociology and chair of the Committee on Education at the University of Chicago, the best use of value-added modeling at present is in evaluating schools and districts, not teachers. At the teacher level, the statistical errors will be more pronounced, he believes, and there are simply too many variables. “We will never get a pure estimate of the efficacy of schools and teachers,” he suggests. “What we will get is that School A might be better than School B, but not because of teachers.” Nevertheless, Raudenbush sees advantages in both value-added assessment and standardized testing. “Accountability is a good thing,” he offers. “But it’s far from a panacea. That’s a big problem in education—too much beating up of teachers and not enough telling them [how they can do better].”

Many union leaders also express interest in value-added modeling, but only if used with caution, and not as a weapon to wield against individuals. “It could be useful,” says Maureen Peters, the Denver union official. “Or it could be used for pitting people against each other.”

Evaluation is Still the Challenge

Value-added modeling seems to be that rare innovation about which even the sharpest critics find much to praise. But the jury remains out. According to its advocates, accurate statistical methods are currently available and are in use. Its opponents point to complications and flaws. They have questions. But they seem willing to see its worth proven—after more research and the perfecting of the statistical techniques involved.

Proponents believe that value-added modeling could return power and credit to the frontline of education: the teachers in the classroom. It could do away with subjective, partisan teacher evaluations. It could,

for the first time, provide a clear way to figure out who are the best, the better, the average and the ineffective teachers. It could offer a way for principals to know which teachers need help and which have the skills to help others.

The value-added method may also hold the potential of becoming a powerful tool in improving struggling low-income, minority schools where it is difficult to attract the best teachers. The efforts of those who do teach in schools where the challenges are great often go unnoticed and unrewarded, but value-added data could provide a way of changing all that. Still, experts warn, struggling schools must be held fully accountable for their students’ educational attainment. Value-added modeling must not be used to lower the bar. Schools with good progress scores but low standardized test scores must still be pushed to improve them.

And there are other caveats.

“I am not one of those who think it’s a silver bullet,” says Adam Urbansky, the Rochester teachers union official. Many education experts sound the same cautionary note. The danger, they say, comes from people in power—ranging from principals to politicians who think value-added modeling is a quick fix. It certainly isn’t that: to begin with, value-added modeling is a complex method and attempts to simplify it—to save money, for example—will only result in inaccurate data and erroneous conclusions. There is also danger in placing too much emphasis on the value-added scores of teachers when, as many continue to point out, student achievement rates may also be the result of socioeconomic and other factors beyond a teachers’ control. And even the most die-hard advocates agree that value-added modeling is useless if it is not acted on. Steps may range from finding help for teachers and schools that need it to more widespread,

imaginative efforts, like community fundraising for incentives for good teachers in troubled schools.

But across the spectrum of opinions about value-added modeling, one common question still begs to be answered: what exactly are the most effective teachers doing that is working so well? If identifying the best teachers is complex and controversial, the process of identifying what they are doing promises to be even more so. Of course, there have long been a variety of theories about what makes a good teacher. But if value-added modeling does, indeed, help to pinpoint the best teachers—either now or later, after more research and refinements—the process of determining what they are doing right may become more reliable, too.

Pamala Carter, a former Hamilton County, Tenn, teacher, now a doctoral candidate at the University of North Carolina in Chapel Hill, is doing research into teaching techniques. Using value-added data provided by Sanders, she is videotaping and studying the methods of teachers in Chattanooga/Hamilton County who have been identified as especially effective. However, it is easier to find these teachers than to figure out what they are doing. “Often they don’t know themselves,” Carter remarks. Her findings so far indicate that the best teachers are adaptable, very well organized and have high expectations for their students. They know their material so thoroughly that

they can easily teach different levels of students. They are flexible enough to use different teaching strategies—to do what is needed to get the material across. And, says Carter, “They’re caring, compassionate, love children and love working with children.”

But those teacher qualities are fairly sweeping. Can schools of education teach aspiring teachers adaptability, flexibility and the love of their material and of working with children? That is a huge challenge. By pinpointing the most effective teachers and schools, value-added modeling could provide the beginning of a new and crucially important area of knowledge.

If, that is, the research into value-added modeling continues. If it is not oversold by politicians and advocates or put in place too rapidly or carelessly—and subsequently discarded by education officials and the same politicians when it does not provide an easy solution to the highly complex issue of student achievement.

“Part of the problem is that the people pushing for it are enthusiasts,” says Raudenbush, the University of Chicago sociologist, which may lead proponents to move too precipitously and thus make mistakes in constructing a truly useful, accurate and effective system of value added modeling. “But,” he adds, sounding more hopeful, “if we’re cautious and continue to do what we all too often don’t do in this country—evaluate—then value-added might really help.”

