

# THE Business Model

*Value-added analysis is a crucial tool in the accountability toolbox—despite its flaws*

by JAY P. GREENE

LIKE THE MAKERS OF HOT DOGS, PSYCHOMETRICIANS, economists, and other testing experts know too well what goes into the creation of achievement tests. Their intimate knowledge of the technical difficulties involved in measuring student achievement makes a number of these testing experts some of the most vocal (and persuasive) opponents of testing. But the flaws in techniques like value-added assessment do not automatically lead to the conclusion that those techniques shouldn't be used to hold educators accountable. Testing may be imperfect, but the alternative—the old system, which allowed us to know very little about the performance of educators—is far, far worse.

To be sure, many of the technical criticisms of value-added testing are correct. It's true that there is more random error in measuring gains in test scores than in measuring the level of test scores. It's true that there is some uncertainty as to whether gains in one area of the test scale are equal to gains at another point in the test scale. And it's true that factors besides the quality of the schools can influence the gains that students achieve. But, on balance, these downsides hardly outweigh the benefits to be reaped from being able to measure and reward productivity in education.

Consider what is likely to continue to happen in education without high-stakes value-added assessment. Unless productivity is measured, however imperfectly, it is not possible to reward teachers, administrators, and schools that contribute most to student learning. If we do not reward productivity, we are unlikely to encourage it. If we do not encourage it, we should not expect more of it.

In fact, this is precisely what has been happening in U.S. education during the past few decades. Between 1961 and 2000, spending on education tripled after adjusting for inflation, from \$2,360 to \$7,086 per pupil. During that time, student performance, as measured by scores on the National Assessment of Educational Progress (NAEP) and high-school graduation rates, has remained basically unchanged. Whenever spending triples without any significant improvement in outcomes, there



is a serious productivity crisis. Yet U.S. public schools just keep chugging along, resisting serious attempts at reform.

Meanwhile, private firms in the United States have been able to achieve steady gains in productivity because the discipline of competition has forced them to adopt systems for measuring and rewarding productivity. Firms that fail to measure and reward productivity lose out to their competitors who do.

Moreover, the systems that private companies use to measure and reward productivity are far from flawless. In fact, the challenge of measuring productivity in the private sector is often

as great as or greater than in education. Imagine a soft-drink company that wishes to measure and reward the productivity of its sales force. The company might determine bonuses (and even decisions on layoffs) based on its salespeople's success at increasing soda sales in their sales area. Like measuring gains in test scores, measuring increases in soda sales is fraught with potential error. Changes in soda sales could be influenced by a variety of factors other than the sales acumen of an employee. Unusually cold weather in an area, a local economic downturn, or exceptional promotional efforts by competitors could all suppress the soda sales of even a very good salesperson. If data on sales are collected using survey techniques, there is also the possibility of random error attributable to the survey method, just as testing has random error. Moreover, if we are comparing sales increases across geographic areas, it is unclear whether it takes more skill to sell soda in an area where the market is already saturated than in an area that initially consumes less soda.

In short, many of the same technical flaws that critics find in value-added testing also exist in the measurement of increases in soda sales. Changes in outcomes may be attributable to factors other than the efforts of the employee. There is random error in collecting the data. And the effort required to produce gains at one level may not be the same as at another level. The only difference is that private firms have rightly not let their inability to achieve the best deter them from pursuing the good.

In the private sector, companies have realized that even flawed evaluation systems nevertheless encourage improvements in productivity. This is because employees cannot be sure that a flawed system will completely obscure the picture of how hard they're working. Employees therefore act as if their productivity were being measured accurately; the chance that slacking will be detected inspires employees to avoid slacking. In fact, evaluation systems with a fairly large amount of error in measuring productivity can still be effective at motivating improvement—if the errors are mostly random, or at the very least do not create perverse incentives, such as encouraging teachers to focus on improving the achievement of one group of students to the exclusion of others.

None of the technical concerns with value-added testing involve perverse incentives. For the most part, the criticisms have to do with random noise in measuring gain scores. Even the nonrandom errors that worry testing critics, such as unevenness in the testing scale or the possible influence of factors outside the school's control, do not create perverse incentives because there are no strong theories about the kinds of behaviors those errors would encourage.

If no one knows what is being mistakenly rewarded, no one has an incentive to engage in that perverse behavior. As long as educators are aware of what the value-added system is supposed to be rewarding, and as long as that system rewards the desired outcomes more than it erroneously rewards something else, the system will help to elicit more of the desired

outcomes—namely, improvements in student achievement.

## The Uses of Data

The development of even an imperfect value-added testing system would revolutionize the systems for hiring, promoting, and compensating teachers. Our current methods provide teachers with little incentive to improve achievement. Promotions and salary increases are based on teachers' seniority and their acquisition of advanced degrees. These characteristics are, at best, weakly related to student achievement. Excellent teachers who possess a master's degree and a few years' experience receive exactly the same salary as lousy teachers with the same formal credentials. Under the current system, we have turned the keys over to educators and trusted that their professionalism will yield improvements in student achievement. Education's productivity crisis in the past four decades should be evidence enough that simple trust is not sufficient.

The development of value-added assessment would also revolutionize how we govern schools and hold them accountable. We currently have little rational basis for saying that a particular school is a good school or that a particular superintendent is a good superintendent. Value-added testing would at least give voters some idea of whether they are getting their tax money's worth out of the school system by giving them at least some information on how the schools are

## In the private sector, companies have realized that even flawed evaluation systems nevertheless ENCOURAGE IMPROVEMENTS IN PRODUCTIVITY.

doing. The fact that voters would have better information on achievement provides the school board with incentives to hire and retain a superintendent who can elicit improvement in student learning. The superintendent, in turn, has an incentive to hire and retain principals who will use the value-added results to hire and promote the best teachers.

Critics of value-added assessment don't necessarily object to using value-added assessment. They object to using the data gleaned from it for high-stakes purposes, such as rewarding or punishing individual schools and teachers. Instead, they suggest that value-added results be provided to administrators so that they can make informed decisions about their employees. This is, to some extent, what happens in the private sector; most pri-

vate firms do not use the crude techniques exemplified by the real-estate company in David Mamet's *Glengarry Glen Ross*, where the salesperson with the fewest sales was fired. Most companies use productivity measures to inform the subjective assessments of supervisors, with some companies permitting less subjective judgment than others for fear of bias or favoritism.

But here is where the parallel between measuring productivity in public education and private industry ends. Supervisors in private companies have incentives to use the information provided by productivity measures properly, because their companies face the discipline of competition from other companies. If supervisors fail to put data on costs, sales, and revenue to good use, their companies will lose out to competitors who do.

In public education, by contrast, local decisionmakers have few or no incentives to make good use of data in assessing their employees because public schools face no meaningful competition. There are basically no consequences for principals who disregard the results of value-added assessments in making decisions about employees, and they're more likely to disregard those results if they consider value-added assessment an unreliable analytical technique. Superintendents will not be able to judge whether principals have used their discretion properly, because they will be told that the value-added test results are not proper grounds for assessing the decisions of principals. And school boards and voters, in turn, will all be stymied in making independent judgments because they will be told that the

## Under the current system, excellent, hard-working teachers who put in tons of overtime receive the same salary as mediocre teachers. WHAT'S FAIR ABOUT THAT?

professional decisions of educators are more reliable than value-added test results.

So there is good reason to fear that principals in public schools, if given discretion to reward teachers as they please, will base their decisions on personal relationships rather than on the results of value-added assessments. In both private industry and public education, a balance must be struck between the mechanical use of productivity measures in assessing employees and relying on the subjective judgments of supervisors. But, in public education, the balance needs to tilt more toward the mechanical application of results, because supervisors have fewer incen-

tives to make appropriate subjective judgments. This means that value-added assessment needs to be high stakes to have the desired positive effect on student learning.

Injustices are unavoidable if value-added assessments are used more mechanically. Some educators will be improperly punished for eliciting what appear to be low gains because of measurement error. Conversely, some educators will be rewarded for improvements for which they were not actually responsible. That said, some of the flaws in value-added assessment have potential technical solutions. For example, if judging a teacher based on his classroom's test scores contains too much error because the sample is too small, we might decide to rate teachers based on a moving average of multiple years of results, thereby increasing the sample size and reducing the random error. But even with technical fixes, some injustices will still occur.

This is not a good reason to abandon the idea, however. After all, some educators will be treated unjustly under any evaluation system. Under the current system, excellent, hard-working teachers who put in tons of overtime receive the same salary as mediocre teachers. What's fair about that? The Buffalo, New York, district recently announced layoffs as a result of a budget deficit. Who were the first to receive pink slips? Not the worst teachers in the district, but the most recently hired. Surely some excellent teachers lost their jobs, while the district retained its burned-out veterans. A peer- or supervisor-review system may reward teachers who are popular among their peers rather than effective with students. Attempting to measure and reward successful educators, with all of its imperfections, is likely to create fewer injustices than any other arrangement. At least using value-added assessments increases the chance that good teachers are rewarded and bad teachers are sanctioned.

Besides, ensuring that every single educator receives justice is at most a secondary concern. An obvious but infrequently recognized truth is that the primary purpose of the education system is to provide a quality education to all students. If high-stakes value-added assessment can help motivate public schools to provide students with a better education, it's a promising reform even if it has some cost to school employees. In no other industry would we even entertain the notion that the interests of employees trump those of customers. Only the political dominance of teacher unions makes us consider the question. It is true that customers usually receive the best service when employees are treated well and fairly. Happily, high-stakes value-added assessment is likely to achieve both ends.

The productivity crisis in public education has certainly created injustices for students and taxpayers during the past few decades. Students have failed to receive high-quality instruction while taxpayers have been paying more and more in return for stagnant test scores. Again, what's fair about that?

—Jay P. Greene is a senior fellow at the Manhattan Institute for Policy Research.