

Published: May 5, 1999

Sanders 101

By Jeff Archer

Knoxville, Tenn.

William L. Sanders has always leaned more toward the research side of academia. In fact, it's been seven years since the statistician last taught a course here at the University of Tennessee.

And yet, in recent months, the balance of his work has tipped heavily toward instruction, and the demand for his teaching has spread far beyond this hilly campus in East Tennessee.

Instead of the graduate students who used to make up his statistics classes, the 58-year-old professor is increasingly being called on to lecture state school boards, legislators, governors, teacher groups, and district officials.

In the past month alone, he has packed up his overheads and jetted off to Florida, Ohio, Delaware, and Kansas. "Sanders 101" has become a very popular course.

It's hardly surprising that his work has struck such a chord. By carefully tracking student progress over time, Mr. Sanders claims to have come up with a way to gauge academic performance—and educators' effects on that performance—that is more accurate and fair than earlier measures.

Mr. Sanders claims to have come up with a way to gauge academic performance—and educators' effects on that performance—that is more accurate and fair

In the current craze for accountability in education, that's like inventing a state-of-the-art mining tool during a gold rush.

"We've absolutely seen an interest from the states in the whole discussion he's started," says Kathy Christie, a policy analyst for the Denver-based Education Commission of the States. "Policymakers are looking for the most viable leverage points."

Closer to home, Volunteer State lawmakers were so impressed with Mr. Sanders' presentation in 1992 that they made his Tennessee Value-Added Assessment System a centerpiece of their school improvement efforts.

Though a visitor may distract him from further research, Mr. Sanders doesn't mind taking time to explain his work. A personable, natural storyteller with an easy drawl, he's been forced in recent months to perfect the arts of analogy and metaphor.

To be sure, his work is complicated, and even if people can't grasp its alphabet-soup equations, he wants them to know the concepts on which it's built. But he also knows that if his methods are misunderstood, people will distrust them. Like the splitting of the atom, his work is seen by some as useful, but potentially dangerous.

“I don’t try to convince everybody,” Mr. Sanders says. “What I try to do with any group is to appeal to the thinkers, because I’m betting that they will become the conduits.”

Ever So Humble

Given the attention Mr. Sanders is drawing, two things are striking about his Value-Added Research and Assessment Center here.

One is its unlikely location in an administrative office at the University of Tennessee’s Institute of Agriculture. The brick building sits on a hill near the Tennessee River, cut off from the main campus by a highway. Along one wall inside, an Agricultural Hall of Fame honors the founders of the country’s “first cooperative livestock marketing association,” and the man who discovered the “hog cholera serum.”

The second is the less-than-plush interior of the center itself, which Mr. Sanders affectionately calls “our little shoe shop.”

The walls, made of an artificial walnut paneling, don’t quite reach the ceiling, eliminating the need for an intercom. The furniture—much of it gray metal and green plastic—dates from the early Space Program period.

Atmosphere aside, though, many agree there’s nothing shabby about the work that’s produced here.

“Sanders himself is a top-notch statistician, and he has really good people working with him,” says Richard Wolfe, who heads the computing department at the Ontario Institute for Studies in Education at the University of Toronto. Mr. Wolfe helped carry out an external review of TVAAS, pronounced “tee-vass,” for the Tennessee comptroller’s office in 1996.

Says Mr. Wolfe: “Statistically, I think the thing he’s come up with is impeccable.”

It has also yielded some striking findings. Mr. Sanders has concluded that teachers are the single most important influence on student progress, an even greater determining factor than socioeconomic status. Simply put, he has found that the effects of a bad teacher, or two consecutive bad teachers, can stick with a child for years. Conversely, the influence of a good teacher can still be seen years down the road.

At the same time, he’s found that highly effective teachers can push students to make significant gains, regardless of their schools’ location.

Some experts see such findings as an important departure. For decades, research on the factors influencing educational success built on the work of the late sociologist James S. Coleman, who began examining such effects in the 1960s.

Though Mr. Coleman’s work has been interpreted in different ways, “the general message taken from Coleman’s findings is that socioeconomic status largely determines student achievement,” says Kati Haycock, who directs the Education Trust, a school improvement effort based in Washington. “And that what schools do doesn’t matter very much, because in the end poor kids learn very little, and rich kids learn a lot.”

Gaining Control

Mr. Sanders arrived at nearly the opposite conclusion by using a remarkably simple concept. While other researchers have spent years struggling to control for differences in students' backgrounds—such as family income and parents' educational levels—Mr. Sanders lets each student act as his or her own control.

To do that, he focuses on gains, instead of on raw scores, so that each student's performance is compared not with that of similar students, but against his or her own past performance.

The idea is that—even if all students don't achieve at the same levels—schools and teachers should at least be adding “value” to each student's performance. Hence the borrowing of the term “value added” from manufacturing, in which each stage of the production process is said to add value to the raw materials.

This approach makes for a fairer evaluation system, Mr. Sanders argues, than one that expects every student to pass over the same bar at the same time.

“I believe that school districts, schools, and individual teachers should never be held for solving all of society's problems,” he says. “But I believe equally strongly that the educational community is responsible for taking each kid as they find that kid and allowing each student, each year, to make academic progress from where he or she is.”

The value-added approach also gives schools with inadequate resources a chance to demonstrate success, even if they don't ultimately perform as well as other systems. Conversely, it exposes schools that are satisfied to rest on their laurels—what Mr. Sanders calls “slide and glide” schools.

Though the concept is easily explained, putting it into practice in a way that attributes gains to individual schools and teachers presents major challenges.

First, it means testing students in each grade, in each subject, every year. Many states only test students at certain levels, such as grades 4, 8, and 10. Fortunately for Mr. Sanders, Tennessee started assessing public school students in grades 2-8 in 1990, and has since only made minor changes in the test instruments. High school tests are now beginning to be used.

The second challenge is reality. Students transfer in and out of districts, educators take breaks for family or medical leave, and teachers do things like team-teach—all of which make it harder to determine who's responsible for what. And even without those complications, Mr. Sanders says, educators simply don't teach enough students each year to yield accurate predictions of future performance—at least not using traditional statistics.

It's in describing how he deals with this problem that Mr. Sanders must stretch his powers of explanation. The tool he uses is called mixed-model methodology. Though written into the Tennessee school code, its exact operation is nearly incomprehensible to a layperson. Mr. Sanders is happy if people grasp what it does.

As he explains it, if a teacher taught just one student for one year and that student made poor progress, then traditional statistics would predict that the teacher's next student would falter as well. Based on only one past result, however, the possibility of error in that prediction would be huge.

Built-In Protection

But mixed-model methodology would take that single result and predict that the next student would make gains that were only slightly worse than the average for all teachers' students. In effect, it's a weighting of results based on how much information one has. The magic actually performed is called "shrinkage estimation," and what it yields is termed a Best Linear Unbiased Predictor, or BLUP.

"What we've attempted to do," says Mr. Sanders, pointing to the low end of a growth curve he has just drawn, "is engineer a process to eliminate, as much as humanly possible, the possibility of someone being put down here just because of some freakishness."

The price he pays is that the estimates of high gains are similarly conservative, he says.

With so many calculations, and so much cross-referencing of data on schools, students, and teachers, the process creates what Mr. Sanders calls "a computing hog," but one that is manageable with software his team has written. Not surprisingly, his center boasts one of the most powerful computers of any department at the University of Tennessee.

"I have told people I've baked cakes with far more complicated recipes," he says. "I've just never had to use a cake pan this big."

For someone on the cutting edge of research on schooling, most of Mr. Sanders' career has been spent outside education.

After earning a doctorate in biostatistics and quantitative genetics, in 1968 he went to work at the federal government's Oak Ridge National Laboratory, the once-secret facility that refined uranium for early atomic weapons. By the time he got there, much of the lab's research had turned toward the effects of radiation on living organisms, and statisticians were in high demand.

In 1972, he moved back to the University of Tennessee, his alma mater, where he began running a center that did statistical analysis for agricultural-research scientists. Says Sanders, "Those people were constantly trying for better ways to take performance data, and to better partition genetic influences from environmental influences, such that they could improve the breeding efficiencies of plants and animals."

So agriculture, as he explains it, is a breeding ground for statistical innovation. A joke in the field goes, "Statistics was raised and reared on the farm." Still, when Mr. Sanders started claiming to see parallels in education, critics sometimes lampooned his background.

But "all science borrows from other sciences," says Willis D. Hawley, a former director of the center for education policy at Vanderbilt University in Nashville.

"Most of the statistics that are used in education are derived from other fields, like sociology or economics," says Mr. Hawley, who now directs the National Partnership for Excellence and Accountability in Teaching, a Washington-based group working to improve teacher quality. "So the fact that he's an agricultural statistician doesn't seem to be relevant. The question is: Is the methodology sensible? And it seems to make sense to me."

Mr. Sanders' first attempts to prove his work's utility, however, went nowhere. In the early 1980s, he learned that then-Gov. Lamar Alexander of Tennessee was seeking a way to award merit pay for teachers. When Mr. Sanders heard someone say it was impossible to do that fairly, he accepted the challenge.

Working with another UT statistician, he completed his first analysis using mixed-model methodology in 1984, drawing on three years of test data from the Knox County schools. The study yielded estimates of teacher effectiveness that were relatively consistent from year to year, and school administrators confirmed that the data jibed well with their own impressions of which teachers were most effective.

But the results failed to spark much interest, Mr. Sanders recalls, and a value-added component never made it into Mr. Alexander's reforms.

Consequences

In 1992, it was a different story. Prodded by a state supreme court order to make Tennessee's school finance system more equitable, state policymakers were looking for ways to increase funding while also holding educators accountable for results.

This time, the legislature bought into the Sanders approach, and his value-added assessment system became an integral part of Tennessee's Educational Improvement Act that year.

As a result, the state now annually releases to the public the most recent three-year-average gains made by each school and district on the Tennessee Comprehensive Assessment Program, or TCAP, the state's nationally normed assessment.

State intervention can be triggered if district- or school-level gains persistently fail to meet the national gains. So far, one district has been placed on probation; the law allows for dissolving local school boards as a last resort.

One of the most powerful and controversial aspects of Mr. Sanders' system is that it can reach beyond the school level to produce a measure of an individual teacher's effectiveness, based on how well the students in his or her classroom perform each year.

In Tennessee, such information is shown only to school officials and to the teachers themselves. State law allows administrators to use the scores in teachers' formal job evaluations after three years of data have been collected. Naturally, that makes some educators edgy.

"You can't walk into a hospital and look at the mortality rates of heart patients and tell who's an excellent physician," says Al Mance, the executive director of the Tennessee Education Association, a National Education Association affiliate. "Sometimes, the best physicians also get most of the patients who are the worst off. So you have to look at each physician's practice, at his or her skills. Human systems are very difficult to pigeonhole."

Mr. Sanders counters that he may not know why certain teachers are more effective than others, but his

One of the most powerful and controversial aspects of Mr. Sanders' system is that it can reach beyond the school level to produce a measure of an individual teacher's effectiveness.

system identifies which ones are getting results. And lest anyone doubt the importance of knowing teachers' effectiveness, he points to research he has carried out showing the lingering impact ineffective teachers have on student performance.

In 1996, Mr. Sanders used data from two Tennessee districts and divided their teachers into five groups—from least effective to most effective. As in TVAAS, each teacher's effectiveness rating was based on students' gains. The researcher found that, on average, students who had been taught by three of the least effective teachers in a row scored below the 50th percentile in mathematics by the end of the third year. By contrast, those who had had three highly effective teachers scored above the 80th percentile.

“The evidence is overwhelming that the percentage of teachers that are just ridiculously ineffective is much smaller than people think—I would judge that to be maybe 3 to 5 percent,” Mr. Sanders says. “But I feel very strongly that even *those* folks should be given some assistance and some time to become more effective, or they should be encouraged to seek employment elsewhere. Because they're harming kids.”

‘Shed Patterns’

Teacher quality aside, some educational researchers are coming to see Mr. Sanders' methodology as a potentially powerful tool for identifying effective instructional techniques, such as block scheduling and cooperative learning. Mr. Sanders himself says he is less interested in weeding out bad teachers than he is in helping teachers who—though not at the bottom—aren't getting the results they want. And he's got a hypothesis on what the problem often is.

In many urban schools, he has noticed a pattern in which students with the lowest past performance make the greatest gains, but those who start with higher scores make little headway. A graph of such gains against past performance creates a downward sloping line from left to right. He calls these “shed patterns,” referring to the sloped roof of a tool shed.

And what they tell him is that a school is focusing on the students at the bottom, while failing to push other children. That emphasis, Mr. Sanders notes, won't show up in overall performance averages for all the students in a school.

That has implications for state accountability efforts. A criticism of some state systems is that by emphasizing a school's percentage of children achieving at or above a certain level, they encourage schools to focus too heavily on certain students at the expense of others.

“It's a lead-pipe cinch that the pace of instruction is going toward the kids with the lowest previous achievement levels, and that the other kids are being held back,” Mr. Sanders says. “In the inner-city schools, a black kid has got a much better chance of being in this pattern than the general population.”

No more desirable, however, is a steep “reverse shed pattern,” suggesting a disproportionate amount of energy being put into teaching the highest achievers. Mr. Sanders would be happy if all students from all levels were at least making gains equal to the national-norm gains—a pattern that would show something close to a horizontal line.

But choosing the right pattern raises fundamental questions about what the ultimate goal of education should be. This is where Ms. Haycock of the Education Trust warns that Mr. Sanders' work may be dangerous.

“He argues that the goal of a teacher, a school, or a system, is that each kid should gain approximately the same amount, and that anything else is unethical,” she says. “I think *that’s* unethical—to say we will take kids from where they are, and that if at the end there’s still a gap, that’s OK.”

Which isn’t to say, she adds, that there isn’t some benefit in a value-added approach. “The interesting question from all of this is, can we marry the idea of standards with the idea that each kid should make gains each year into something that doesn’t lose the advantages of each?” Ms. Haycock says.

In the 2001-2 school year, in fact, Tennessee plans to begin requiring all high school students to pass “gate keeper” tests before receiving a state-approved diploma. In the meantime, many are betting that Mr. Sanders’ system will push schools and teachers enough to ensure that students will be able to pass over that bar.

“When you get to the bottom line and passed all the mixed metaphors,” he says, “it’s about giving more kids more opportunities in life. That should be the target for all of us.”

Vol. 18, Issue 34, Pages 26-28