latimes.com

'Value-added' teacher evaluations: L.A. Unified tackles a tough formula

Los Angeles school district leaders are poised to plunge ahead with their own confidential 'value-added' ratings this spring, saying the approach is far more objective and accurate than any other evaluation tool available, despite its complexity.

By Teresa Watanabe, Los Angeles Times

March 28, 2011

In Houston, school district officials introduced a test score-based evaluation system to determine teacher bonuses, then — in the face of massive protests — jettisoned the formula after one year to devise a better one.

In New York, teachers union officials are fighting the public release of ratings for more than 12,000 teachers, arguing that the estimates can be drastically wrong.

Despite such controversies, Los Angeles school district leaders are poised to plunge ahead with their own confidential "value-added" ratings this spring, saying the approach is far more objective and accurate than any other evaluation tool available.

"We are not questing for perfect," said L.A. Unified's incoming Supt. John Deasy. "We are questing for much better."

As value-added analysis is adopted — if not embraced — across the country, much of the debate has focused on its underlying mathematical formulas and their daunting complexity.

All value-added methods aim to estimate a teacher's effectiveness in raising students' standardized test scores. But there is no universal agreement on which formula can most accurately isolate a teacher's influence from other factors that affect student learning — and different formulas produce different results.

Nor is there widespread agreement about how much the resulting ratings should count. Tensions are all the greater because the stakes for teachers are high as more districts consider using the evolving science as a factor in hiring, firing, promotions, tenure and pay.

"It is too unreliable when you're talking about messing with someone's career," said Gayle Fallon, president of the Houston Federation of Teachers.

She said many teachers don't understand the calculations. The general formula for the "linear mixed model" used in her district is a string of symbols and letters more than 80 characters long:

$y = X\beta + Zv + \varepsilon$ where β is a p-by-1 vector of fixed effects; X is an n-by-p matrix; v is a q-by-1 vector of random effects; Z is an n-by-q matrix; $E(v) = 0$, $Var(v) = G$; $E(\varepsilon) = 0$, $Var(\varepsilon) = R$; $Cov(v,\varepsilon) = 0$. $V = Var(y) = Var(y - X\beta) = Var(Zv + \varepsilon) = ZGZT + R$.

"It's doctorate-level math," Fallon said.

In essence, value-added analysis involves looking at each student's past test scores to predict future scores. The difference between the prediction and students' actual scores each year is the estimated "value" that the teacher added — or subtracted.

The Times released a value-added analysis of about 6,000 L.A. Unified elementary school teachers in August that was based on district data. Before school ends, L.A. Unified plans to release its own analysis, confidentially providing teachers with their individual value-added scores. For at least the first year, the teachers' scores will not be used in formal evaluations; whether they are ultimately used is subject to negotiation with the union.

Deasy and many others argue that value-added analysis is far more useful than the common practice of dispatching administrators to classrooms, where they often make pro-forma observations. These reviews overwhelmingly result in "satisfactory' ratings, which may or may not be deserved.

In designing its model, the nation's second-largest school district has wrestled with myriad questions: whether to tweak the model to account for the students in a class who don't speak fluent English, for example, or for those who moved from one school to another during the academic year.

Should value-added models take student race and poverty into account, even if it means having lower expectations for some races and higher ones for others?

Deasy said these were among the most difficult questions the district grappled with. Theoretically, value-added models inherently account for these differences, because each student's performance is compared each year with the same student's performance in the past, not with the work of other students. But many experts say further statistical adjustments are necessary to improve accuracy.

A 2010 study of 3,500 students and 250 teachers in six Bay Area high schools by researchers at Stanford University and UC Berkeley found that, under their model, teachers with more African American and Latino students tended to receive lower value-added scores than those with more Asian students.

Dan Goldhaber, a professor at the University of Washington Bothell, said that there is no definitive answer on the race question but that most specialists in the field support factoring it in because research overwhelmingly shows that it is correlated with student performance.

William Sanders, value-added consultant for the Houston Independent School District, strongly opposes adjusting for race or socioeconomic status, however. He says that it is unnecessary and that adjustments would camouflage such institutional problems as the inequitable distribution of teaching talent. "I want administrators to deal with this and not sweep it under the rug," he said.

Deasy said that after long internal debate, L.A. Unified decided to control for race, ethnicity, mobility, English proficiency and special education status. He noted that they can affect achievement but "don't determine or predict it."

All these decisions about what to account for in the formula can add to its complexity and affect results.

In an analysis of The Times' teacher rankings last month, Derek Briggs and Ben Domingue, researchers at the University of Colorado at Boulder, began with a similar data set but factored in more variables than used in the Times model, such as the effect of a student's peers on academic performance. They found their changes produced different effectiveness ratings for 54% of teachers in reading and 39% of teachers in math.

"If you change a model in this way and get different results, it raises questions about its accuracy," said Briggs, an associate professor of research evaluation methodology.

Others disagreed. The consultant who prepared The Times analysis, economist Richard Buddin, said the discrepancies that pushed teachers into different categories — for example, from "average" to "more effective" — often were very small. That points to the problem inherent in grouping teachers into categories, not to a fundamental flaw in the analysis.

Even if the value-added method is sound, its application can be flawed. In New York, for instance, several teachers found multiple errors in their value-added reports — teachers scored for students they had not taught or math teachers getting English reports, according to the United Federation of Teachers.

Administrators agree that such errors are unacceptable and say they constantly adjust their methods. The Hillsborough County School District in Florida, for example, has developed safeguards to ensure that teachers are matched with the students they actually teach.

Many teachers and union leaders say they are not necessarily opposed to value-added methods but want to understand them and have a say in how they're used.

Sarah Bax, an eighth-grade math teacher in Washington, D.C., started sending emails to district administrators last September, taking issue with her value-added scores and asking for the algorithm used. After five months, she was told that no written information would be available until May.

"How do you justify evaluating people by a measure [for] which you are unable to provide explanation?" she wrote to a local school official in an email.

In Los Angeles, outgoing United Teachers Los Angeles President A.J. Duffy said he's fine with using the method to offer teachers information but opposes including it in performance reviews and other high-stakes decisions.

"It is off the table as far as we're concerned," Duffy said. "The more complicated a statistical approach to analyzing human behavior is, the more likely it will rely on generalities that are wildly inaccurate."

Nathan A. Saunders, president of the Washington Teachers Union, said he sees a role for value-added data as one of many components in teacher evaluations, given what he called reasonable weight, 20% or less, compared to 50% in the District of Columbia. (Virtually no one supports using value-added as a sole measure of performance.)

"Fifty percent is enough to sink the ship," he said.

As districts choose their models, the concern will only rise.

"The political heat is fixin' to get huge," Sanders said.

teresa.watanabe@latimes.com