

The High Stakes of Teacher Evaluation

By Jack Schneider

Teacher behavior matters less than student learning. That's the new mantra in education reform. From coast to coast, classroom observations are being replaced by student achievement scores as the coin of the realm in teacher evaluation. In many states, 50 percent of teacher effectiveness ratings are now determined by such data, and if the trend continues that number will only grow. It won't be long before high-stakes personnel decisions—hiring, firing, and pay raises—are conducted by computers running algorithms rather than by administrators toting clipboards.

Teachers and union leaders, for their part, have strongly resisted this shift, channeling their opposition into two primary criticisms, neither of which has been particularly effective.

The first argument is that teacher evaluation reform is a Trojan horse. Its real purpose, they argue, is to undermine job security. As one union rep put it, “they're not focused on improvement; they're focused on kicking people out.” To draw on a poker analogy, reformers are raising the stakes of the game and forcing teachers to go all-in. Win, in other words, or go home.

The second argument is that evaluation schemes based on student achievement data produce inconsistent results from year to year. To stick with the analogy, reformers are appraising poker skill based on the results of a single hand. A bad player, of course, will occasionally win, just as a good player will lose. But that's not skill; that's chance.

These concerns are valid. Yet such arguments have largely failed to curb policymakers or sway the public because reformers have so effectively countered them, portraying teachers as self-interested impediments to reform.

Consider the question of undercutting job security. Reformers make no apologies for the fact that quantification schemes will erode tenure and cultivate a castigatory atmosphere. Why, they ask, should that not be the case? Ineffective educators don't belong in the classroom and our current approaches to evaluation have failed to weed out that entrenched minority. In many cases, highly superficial classroom observations are conducted once per year, and most teachers

receive the highest possible rating. As Indiana assistant superintendent Dale Chu framed it, a quarter of the state's third-graders "are reading or computing at a minimal level," yet "99 percent of teachers ... are rated as effective or highly effective." The status quo, reformers assert, is poker with no stakes.

No wonder, then, that they want to dump the old system and adopt a new one. Student achievement data is objective and uncompromising. Sure, the stakes are high and there will be losers. But, as they tend to point out, there will also be big winners. In Washington, DC under the district's new IMPACT model, highly effective teachers can make up to \$140,000 annually. As the district's head of human capital declared: "we want to make great teachers rich." Good poker players, in other words, don't fear high stakes; they seek them out.

Responding to the second argument, reformers will concede that their data systems tend to be somewhat erratic. Of course, that would be hard not to admit. In study after study, researchers have found that even the most advanced models produce significantly different results from year to year. Teachers from the top 20 percent one year can end up in the bottom 20 percent the next. And according to a recent Mathematica study, the error rate for comparing teacher performance was 35 percent. 65 percent accuracy, in the world of education, generally earns a D.

Not to worry, reformers argue, the numbers eventually smooth out. In one hand of poker anything can happen—luck can bring garbage or it can bring four aces. Over time, however, results regress to the true mean. Talent will out, and bad players will go bust.

In city after city and state after state, this is how the argument goes. Teachers express their concerns and reformers counter. And the end result is that the momentum behind data-driven teacher evaluation schemes continues to build.

But there is another case that teachers might make—a criticism that would level a blow to the radical overhaul of teacher evaluation, and more importantly, one that just might help students learn. And the case is this: achievement, as we measure it, is not really about achievement. As determined by multiple choice tests—the dominant way that we measure it in the U.S.—achievement is not about how students can think or write or persuade. It is not about how they can perform experiments or produce original research. It is not about their prowess in art or civics or robotics. Instead, it is about memorized minutia and good guesses. We accept this

approach to measurement only because it is so common. And it is common not because it actually measures achievement, but because it is time efficient and cost-effective.

Simply put, we're using the wrong instrument. Evaluating teachers through multiple choice-based tests of student learning is like using the rules of Go Fish to assess poker skill. Instead of learning how to evaluate complex hands like flushes, straights, and full houses, we're asking teachers if they have any sevens. It's a much simpler and, ultimately, much less interesting game.

This doesn't mean that we should turn our backs on data or stop trying to gauge teacher quality. It doesn't mean that outstanding teachers need go unrewarded or that their ineffective peers must be protected. Instead, to paraphrase Rich Shavelson, it means we need to take audacious steps to measure a more full set of learning outcomes—outcomes valued by teachers, scholars, and the American public. It means moving beyond multiple choice tests, and developing assessments oriented toward performance and habits of mind.

In the meantime, firing teachers based on deficient measures of effectiveness is a reckless proposition, and educators are right to oppose it. But they need to be savvier about the way they are taking on this fight. Job security is not an argument they can win on. Tenure and seniority, after all, are poor indicators of teacher quality, and in backing the status quo, they allow themselves to be portrayed as barriers to change. The weakness of the statistical measures is also not a winning talking point. The math involved is too complicated for laypeople to decipher, and the aggregate research is easily cherry-picked.

The real issue, and the one that teachers need to take a stand on, is that high-stakes personnel decisions based on Go Fish measurements will have one of two destructive outcomes. The first—a drastic dumbing-down of instruction—has already started to take place as a result of NCLB's crude accountability measures. But when teachers' jobs are on the line, the floodgates will open; overall quality of instruction will decline not by degree, but by orders of magnitude.

The second outcome, equally likely and equally problematic, is a potential exodus of great teachers from the profession. Many, of course, will take their lumps. They will continue teaching students to think, to write, to play the violin, or to take carbon dioxide gas samples; and they may suffer for it. But others will leave. Unwilling to play a thoughtless and artless game,

they will stand up and leave the table. And they won't come back.

Jack Schneider is author of *Excellence for All: How a New Breed of Reformers is Transforming America's Public Schools* (Vanderbilt University Press, 2011). This fall he will begin an appointment as Assistant Professor of Education at Holy Cross in Worcester, Mass.