

EDUCATION WEEK

Published Online: February 5, 2013

Published in Print: February 6, 2013, as **High Ratings For Teachers Are Still Seen**

Teachers' Ratings Still High Despite New Measures

Changes to evaluation systems yield only subtle differences

By **Stephen Sawchuk**

The figures are resoundingly familiar.

In Michigan, 98 percent of teachers were **rated effective or better** under new teacher-evaluation systems recently put in place. In Florida, 97 percent of teachers **were deemed effective or better**.

Principals in Tennessee judged 98 percent of teachers to be "at expectations" or better last school year, while **evaluators in Georgia** gave good reviews to 94 percent of teachers taking part in a pilot evaluation program.

Those results, among the first trickling out from states' newly revamped yardsticks, paint a picture of a K-12 system that remains hesitant to differentiate between the best and the weakest performers—as well as among all those in the middle doing a solid job who still have room to improve.

The data are also raising new questions about the observation components of the systems, which tended to produce the highest scores.

"Most of these districts are trying really hard to think about teacher evaluation in a good way and to use it developmentally, but there's still some cultural challenges," Sarah W. Lenhoff, the assistant director of policy and research at the Education Trust-Midwest, said about the Michigan results. The Royal Oak, Mich.-based advocacy group has issued several reports on the overhaul of evaluation in that state.

"Although teachers may be getting more feedback and talking about their practice more, it hasn't trickled down to variations in rating," Ms. Lenhoff continued. "And that's going to take time."

Some teachers' unions, though, see the data as an affirmation of teachers' hard work.

"Despite all the rhetoric blaming teachers for all the problems in education, most teachers are doing a

[← Back to Story](#)

EDUCATION WEEK Multi-User LICENSING



**K-12 IS
INNOVATING.**
Keep your team
up to speed

+ Click here for more info

good job, given their limited resources," said Doug Pratt, a spokesman for the Michigan Education Association.

While only a handful of states already have a year or more of evaluation data, a host of others are now in the midst of full-scale implementation and will release initial results later this year.

Dozens of states have taken steps in recent years to overhaul their teacher-evaluation systems, often in response to federal incentives. Such changes have also been promoted by an influential lineup of organizations that calls for greater accountability in the teaching profession. The states hope to use the systems to strengthen teaching practices and dismiss poorly performing teachers.

Translating Results

In Florida, where every district was required to implement a new teacher-evaluation system in 2011-12, data released in December show that 97 percent of teachers received one of the top ratings. That figure, while high, is still lower than the 99.9 percent from before the revisions, state officials noted.

"We know that in the first year, most districts exercised an abundance of caution," said Kathy Hebda, the state chancellor for educator quality. "We said upfront that our plan was to start together, and to get better every year. We do think it was a really good start, considering how big we are, and how much work there was to do."

Evaluation data in Georgia, also released in December, have so far been limited to 26 districts participating in the state's federal Race to the Top grant. In 2012, administrators there essentially had to navigate two evaluation systems: the pilot program and the one required by the terms of existing laws and board policies.

"It's not likely that many principals are going to rate teachers differently on the pilot system than the system they're using for [human resources] purposes," said Martha Ann Todd, the state's associate superintendent for teacher and leader effectiveness.

Michigan's high numbers, released in November, could point to uncertainty in that state about the process, which remains somewhat in flux. A council working to outline a model state system for districts has not yet made its recommendations to the legislature.

Scores High

The early results offer several possible interpretations. As scholars have pointed out, there is no consensus about the percentage of teachers who should be identified as underperforming or superior in any given year.

"I do think we are still in a space of trying to do the research ... as these systems are being implemented, making sure that we are following up on things like alignment between the different measures," said Laura

Changes Ahead

More than half of states will be implementing revised teacher-evaluation systems in the next three years.

Some early adopters have already begun to make or propose changes to key aspects, such as the weight given to quantitative measures, based on early results and feedback from teachers.



SOURCE: *Education Week*

Goe, a research scientist in the Learning and Teacher Research Center of the N.J.-based Educational Testing Service.

"We just don't have enough large-scale research studies yet to say that this is the right way to do it at the school level, the district level, or the state level," she said.

One of the key areas of congruence throughout the state data from Florida, Tennessee, and Georgia is the generally high scores given to teachers during classroom observations, a finding that comes right as new research is revealing clues about the properties of such observations and how they are shaped by the norms within schools.

A **2010 study on a pilot system in Chicago** found, for instance, that when using the high end of the scale, principals often inflated their ratings compared to other observers, in part because of cultural expectations. And researchers have also found that having several different people observe teachers **helped make their judgments more consistent** over time and mitigate such bias.

Some districts say they worked deliberately to reach more nuanced findings. Florida's Lee County showed a broader range of scores than most other districts in the state. Under its system, about 1.5 percent of teachers received unsatisfactory ratings, compared with 0.2 percent statewide.

Most Lee County teachers' instruction was deemed solid, but the district reserved the highest category for only the top 10 percent of teachers. (Statewide, 23 percent of teachers earned that "highly effective" rating.)

Officials for the school district say they worked jointly with the local teachers' union to agree on how state-generated test-score data would be interpreted, and took pains to make the evaluation instrument clear, to help make those finer distinctions in performance.

"We're still working on developing training. It's huge, and very important, because our goal is to bring about the real consistency in our schools about how we evaluate," said Greg Adkins, the chief negotiator for the 85,000-student Lee County district.

Multiple Measures

The introduction of quantitative, supposedly objective data also might help ensure a broader range of scores.

"With value-added in particular, you are essentially ranking results for teachers, so ... you have some who are necessarily going to be closer to the bottom. Whereas with observations you can have all the teachers on the top," said Ms. Goe, who also advises the Great Teachers and Leaders Center, a federally funded technical-assistance provider housed at the American Institutes for Research.

"Value added" is a statistical method of estimating the effect of a teacher's instruction on his or her students' test scores.

Tennessee's data released last summer show, for instance, that observers gave only 0.2 percent of teachers the lowest score, compared to quantitative measures that put 16.5 percent of teachers in that category.

Georgia officials are still examining the quantitative portion of the pilot data, but preliminary reports on

"student learning objectives"—district-determined common growth measures—showed more variability than did observations.

Disparities between different measures ought to invite more investigation, Ms. Goe of the ETS said. It could mean that "there is a mismatch between what we can see the teacher doing and how the learning is taking place," she said, "or that there are other factors entering into the situation."

Tennessee has already begun such analyses. This school year, the state sent "coaches" to 73 schools that had high teacher ratings and low levels of student-growth in 2011-12. The goal was to help support and retrain evaluators on how to document teacher practice.

"We've committed to a process of continuous improvement for this evaluation system, and when we saw this need, we acted on it," said Kelli Gauthier, a spokeswoman for the Tennessee education department.

Midyear data suggest that results from those schools are likely to be distributed more broadly, although a majority of teachers will still pass muster.

Georgia officials are also redoubling efforts to improve consistency across raters. So far, the state has trained about 3,400 evaluators in face-to-face or online sessions, Ms. Todd, the associate superintendent, said. Some districts are also creating libraries of teaching videos to help refresh evaluators' memories of exemplary performance at each level. The challenges for ensuring inter-rater reliability are somewhat greater for states that have given more discretion to their districts regarding training.

"I think there's a huge incentive for the state to invest in principal training," said Ms. Lenhoff said of Michigan, "and make sure every principal understands and has practiced how to observe teacher behavior, how to take notes, how to give feedback to teachers."

Coverage of policy efforts to improve the teaching profession is supported by a grant from the Joyce Foundation, at www.joycefdn.org/Programs/Education.

Vol. 32, Issue 20, Pages 1,18-19

RELATED BLOG



[Visit this blog.](#)