

[Education Week's blogs](#) > [Teacher Beat](#)

## Class Composition Can Bias English Teachers' Observation Scores, Study Finds

By [Stephen Sawchuk](#) on January 22, 2016 7:04 AM

Observations of teachers—usually the most prominent component of teacher-evaluation systems—can carry significant sources of bias, potentially penalizing English/language arts teachers of lower-achieving students, [concludes a recent research study](#).

The paper, by Matthew Steinberg of the University of Pennsylvania and Rachel Garrett of the American Institutes of Research, was published in the January edition of the journal *Educational Evaluation and Policy Analysis*.

The analysis provides more evidence that, despite the widespread concern about test-score-based ratings of teachers, observations of teachers are just as susceptible to error. It [deepens earlier findings](#) by looking at the topic at a more granular level, and by showing that the findings are consistent across several analytical samples.

For the study, the researchers examined observation scores of some 834 teachers, who were rated by trained reviewers on Charlotte Danielson's Framework for Teaching. (The data comes from the the Bill & Melinda Gates Foundation's Measures of Effective Teaching study, [which examined several different measures of teacher performance](#).)

The researchers found that ELA teachers of higher-achieving students were more likely to get higher scores on the "classroom environment" domain of the Framework for Teaching, which considers teachers' classroom management and their ability to create a respectful learning environment, among other goals. Overall, an English teacher with a class whose incoming achievement was a standard deviation higher would get his or her score pushed up by about a third of a standard deviation.

The study found that the pattern was present even among a subset of teachers who were randomly assigned a class of students—an important distinction because that reseach technique helps filter out other potential sources of bias.

Higher-performing students are less likely to have behavior infractions, the authors postulate. "It sort of makes sense," Steinberg said. "If you have better behaved students, it's probably easier to work with them and to develop that rapport."

Now here's the interesting thing. The bias did not show up on all of the FFT domains. Teachers' ratings on things directly related to instructional technique—whehter they ask students probing questions and use assessments for instruction—weren't related to students' prior achievement. And math teachers generally didn't seem to get the same boost from having better-performing students. The authors postulate that math instruction, by its very nature, tends to be more direct and to use fewer groups, conversations, and other techniques that rise and fall on student interactions.

What does this mean for policy? Should states and districts chuck out pieces of these observation protocols that seem susceptible to this kind of bias? Steinberg doesn't think so, since the information from these systems can still be used as a tool to help teachers improve overall.

But, if these indicators are going to factor into a rating that might affect teacher pay, promotion, or job security, then some tweaks might be needed—like using a rolling average over several years to help even out swings in performance due to classroom composition.

"Practitioners and policymakers really need to meet in the middle on this," he said.