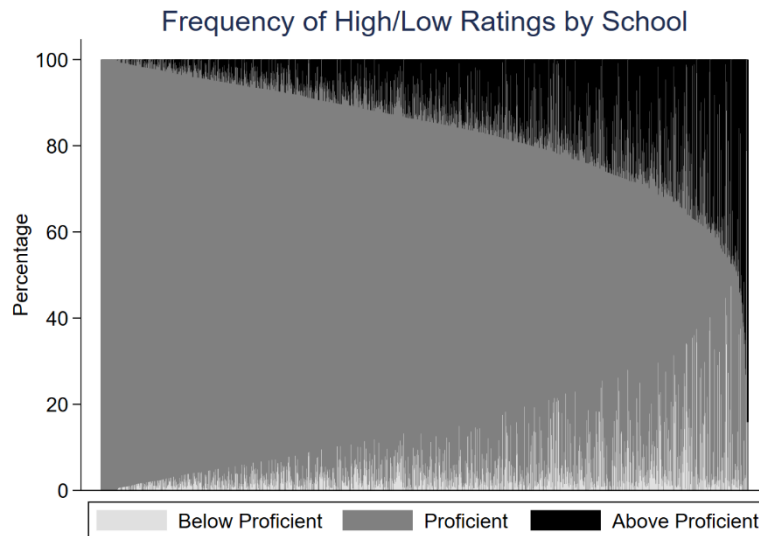


Performance Evaluations as a Measure of Teacher Effectiveness when Standards Differ

James Cowan, Dan Goldhaber, Roddy Theobald

Overview

We use statewide data from Massachusetts to investigate the school role in teacher evaluation. Schools classify most teachers as proficient but differ substantially in how frequently they assign other ratings. We show these patterns are driven by differences in the application of standards across schools, not by differences in the distribution of teacher quality.



Above, we show the distribution of performance ratings across all schools with at least 50 evaluations ($N = 1,610$). Each vertical stripe depicts the percentage of teachers receiving below proficient, proficient, or above proficient ratings in a single school. Although most teachers (85%) receive a proficient rating, schools vary substantially in the extent to which they differentiate between teachers in their performance ratings. Schools on the left side give nearly every teacher a proficient rating, while those on the right side give this rating to only about half of their teachers.

Evidence from Teacher Transfers

To assess whether these patterns reflect variation in the implementation of evaluation policies, we examine teachers' performance ratings before and after they transfer schools. We divide schools into two groups (high and low variance) based on their evaluation histories. High variance schools are those with greater observed variation in evaluations (right side of the previous figure). Compared to teachers who leave high variance schools for other high variance schools, teachers who leave for low variance schools are about 5 percentage points less likely to receive both an above proficient rating *and* a below proficient rating in the following year. These trends are reversed for teachers leaving low variance schools. We see little effect on average ratings or on student performance, which suggests that these findings are caused by variation in evaluation practices.

Implications

We find that teacher performance ratings depend on the school and classroom context. These findings suggest caution when using performance evaluations to make high-stakes comparisons between teachers in different school settings. States that use locally implemented performance evaluations for statewide accountability efforts may also need to ensure that they are consistently implemented across schools and districts. Because our findings are driven in part by the "widget effect" in schools that provide the same rating to most teachers, extracting more useful information from teacher evaluations may require changing the extent to which they differentiate between teachers. Finally, researchers should be aware that evaluations may signal different information about teaching quality in different locations.

Full Article Citation:

Cowan, J., Goldhaber, D., and Theobald, R. (2022). [Performance Evaluations as a Measure of Teacher Effectiveness when Standards Differ](https://doi.org/10.1080/19345747.2021.2018747). *Journal of Research on Educational Effectiveness*. DOI: <https://doi.org/10.1080/19345747.2021.2018747>.