

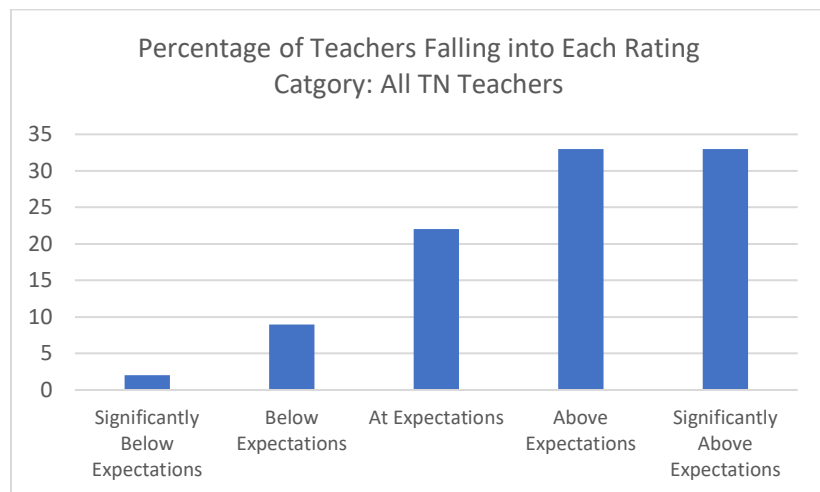
Teacher Performance Ratings and Professional Improvement

Cory Koedel, Jiayi Li, Matthew G. Springer, & Li Tan

Do Rating Differences in Reformed Teacher Evaluation Systems Cause Teachers to Alter Their Professional Improvement Behaviors?

According to our analysis of Tennessee’s reformed teacher evaluation model, the answer is no.

In the past decade, states have focused on developing new teacher evaluation systems that better differentiate teachers by their levels of effectiveness, and promote teacher development. But our findings suggest that in the case of Tennessee, differences in ratings from the new system did not cause teachers to respond by changing their improvement behaviors. When comparing groups of similarly effective teachers that were assigned to lower vs. higher effectiveness categories, we found no meaningful differences in time spent on improving practice or changes in professional improvement strategies. This indicates that being assigned to a lower vs. higher rating category, within Tennessee’s reformed evaluation system, did not cause teachers to change their behaviors along these dimensions.



Notes: These percentages are taken from Table 1, Column 1 of the article.

What does Tennessee Measure, and Why Would it Matter?

Tennessee’s new evaluation system scores teachers on a 500-point scale using student learning growth, student achievement, and traditional qualitative measures like classroom observation. Teachers are assigned to one of five effectiveness categories that are defined by point thresholds. Teachers are distributed among the rating levels in the new Tennessee system more widely than is typical of older evaluation systems (see chart at bottom-left). Thus, the new ratings might have influenced teacher behavior due to the influx of information about one’s relative performance, the psychological effects of category assignment, and informal local policies that reward or sanction teachers based on ratings.

Methods and Setting

Since teacher rating categories were defined by score cutoffs, teachers with nearly identical scores were assigned to different rating categories. Our study uses this, and the fact that teachers know their rating category (e.g., “at expectation”) but not their score (e.g., 357), to measure the behavioral differences between groups of otherwise comparable teachers assigned to different rating categories. This research technique is called a regression discontinuity design, and it allows us to rigorously estimate the causal effect of receiving a lower vs. higher rating.

Teacher behaviors were measured using four items on a state-administered survey asking teachers how much time they spent on professional improvement activities and whether their assessment score changed their teaching methods, professional development activities, or non-teaching activities. Our sample was limited to those with scores near cutoffs and those who responded to the survey. The cutoff comparisons have sample sizes ranging from 3,000-7,700. Future research should explore if different types of feedback received as part of the evaluation system—i.e., outside of the formal ratings—affect teacher behaviors or performance.

Full Article Citation:

Koedel, Cory, Li, Jiayi, Springer, Matthew G. & Tan, Li (2019). [Teacher Performance Ratings and Professional Improvement](#). *Journal of Research on Educational Effectiveness*, 12:1, 90-115, DOI: 10.1080/19345747.2018.1490471