

Teacher peer observation and student test scores: Evidence from a field experiment in English schools

Simon Burgess (*University of Bristol*)

Shenila Rawal (*Oxford Partnership for Education Research and Evaluation*)

Eric Taylor (*Harvard Graduate School of Education*)

May 2018

This paper reports improvements in teacher performance, as measured by student test scores, resulting from a program of teacher peer evaluation. A consistent literature now documents educationally and economically meaningful differences in teacher job performance, especially as measured by their contributions to student test scores. However, the causes of those differences are far from well understood, and few management or policy interventions have been shown to improve the job performance of incumbent teachers (for a recent review see Jackson, Rockoff, and Staiger 2014). We conducted a large-scale randomized experiment in which teachers within a school observed and scored each other's teaching, following a structured rubric (a modified version of Charlotte Danielson's *Framework for Teaching*) and using tablet computers. The intervention lasted two years and covered classes in English and maths for pupils working towards key high-stakes tests. The RCT comprised three nested experiments. First, schools were randomly assigned to implement the peer evaluation program (treatment) or continue business as usual (control). Second, treatment schools were further randomly to either a high dose condition for maths and low dose for English, or high for English and low for maths. In the high dose condition the program suggested twice as many peer observations as the low dose. Third, all teachers in treatment schools were randomly assigned to be an observer, to be observed, or to participate in both roles. Our data include GCSE scores for some 40,000 English students in 82 schools. In preliminary results, we find GCSE scores were approximately 0.06 student standard deviations higher in treatment schools during the experiment years. We also find that, while high dose departments did roughly twice as many peer observations, the treatment effects were not larger. Finally, we find that teacher performance improved both for observers and observees, perhaps more for observers.

Jackson, K., Rockoff, J. and Staiger, D. (2014) Teacher Effects and Teacher-related Policies. *Annual Review of Economics* vol. 6 pp. 801 - 825