Economics of Education Review 48 (2015) 16–29

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement

David Blazar*

Harvard Graduate School of Education, Center for Education Policy Research, 50 Church Street, 4th Floor, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 24 June 2014
Revised 14 May 2015
Accepted 15 May 2015
Available online 27 May 2015

Keywords:
Teacher quality
Instruction
Mathematics education
JEL Classifications: Analysis of Education (I21)
Human Capital (J24)
Econometrics (C01)

ABSTRACT

Recent investigations into the education production function have moved beyond traditional teacher inputs, such as education, certification, and salary, focusing instead on observational measures of teaching practice. However, challenges to identification mean that this work has yet to coalesce around specific instructional dimensions that increase student achievement. I build on this discussion by exploiting within-school, between-grade, and cross-cohort variation in scores from two observation instruments; further, I condition on a uniquely rich set of teacher characteristics, practices, and skills. Findings indicate that inquiry-oriented instruction positively predicts student achievement. Content errors and imprecisions are negatively related, though these estimates are sensitive to the set of covariates included in the model. Two other dimensions of instruction, classroom emotional support and classroom organization, are not related to this outcome. Findings can inform recruitment and development efforts aimed at improving the quality of the teacher workforce.

1. Introduction

Over the past decade, research has confirmed that teachers have substantial impacts on their students' academic and life-long success (e.g., Nye, Konstantopoulos, & Hedges, 2004; Chetty, Friedman, & Rockoff, 2014). Despite concerted efforts to identify characteristics such as experience, education, and certification that might be correlated with effectiveness (for a review, see Wayne & Youngs, 2003), however, the nature of effective teaching still largely remains a black box. Given that the effect of teachers on achievement must occur at least in part through instruction, it is critical that researchers identify the types of classroom practices that matter most to student outcomes. This is especially true as schools and districts work to meet the more rigorous goals for student achievement set by the Common Core State Standards (Porter, McMaken, Hwang, & Yang, 2011), particularly in mathematics (Duncan, 2010; Johnson, 2012; U.S. Department of Education, 2010).

Our limited progress toward understanding the impact of teaching practice on student outcomes stems from two main research challenges. The first barrier is developing appropriate tools to measure the quality of teachers' instruction. Much of the work in this area tends to examine instruction either in laboratory settings or in classrooms over short periods of time (e.g., Anderson, Everston, & Brophy, 1979; Star & Rittle-Johnson, 2009), neither of which is likely to capture the most important kinds of variation in teachers' practices that occur over the course of a school year. The second is a persistent issue in economics of education research of designing studies that support causal inferences (Murnane & Willett, 2011). Non-random sorting of students to teachers (Clotfelter, Ladd, & Vigdor, 2006; Rothstein, 2010) and omitted measures of teachers' skills and practices limit the success of prior research.

I address these challenges through use of a unique dataset on fourth- and fifth-grade teachers and their students from three anonymous school districts on the East Coast of the

* Corresponding author. Tel.: +1 617 549 8909
E-mail address: david_blazar@mail.harvard.edu

http://dx.doi.org/10.1016/j.econedurev.2015.05.005
0272-7757/

United States. Over the course of two school years, the project captured observed measures of teachers' classroom practices on the Mathematical Quality of Instruction (MQI) and Classroom Assessment Scoring System (CLASS) instruments, focusing on mathematics-specific and general teaching practices, respectively. The project also collected data on a range of other teacher characteristics, as well as student outcomes on a low-stakes achievement test that was common across participants.

My identification strategy has two key features that distinguish it from prior work on this topic. First, to account for sorting of students to schools and teachers, I exploit variation in observation scores within schools, across adjacent grades and years. Specifically, I specify models that include school fixed effects and instructional quality scores averaged to the school-grade-year level. This approach assumes that student and teacher assignments are random within schools and across grades or years, which I explore in detail below. Second, to isolate the independent contribution of instructional practices to student achievement, I condition on a uniquely rich set of teacher characteristics, skills, and practices. I expect that there likely are additional factors that are difficult to observe and, thus, are excluded from my data. Therefore, to explore the possible degree of bias in my estimates, I test the sensitivity of results to models that include different sets of covariates. Further, I interpret findings in light of limitations associated with this approach.

Results point to a positive relationship between ambitious or inquiry-oriented mathematics instruction and performance on a low-stakes test of students' math knowledge of roughly 0.10 standard deviations. I also find suggestive evidence for a negative relationship between teachers' mathematical errors and student achievement, though estimates are sensitive to the specific set of teacher characteristics included in the model. I find no relationships between two other dimensions of teaching practice – classroom emotional support and classroom organization – and student achievement. Teachers included in this study have value-added scores calculated from state assessment data similar to those of other fourth- and fifth-grade teachers in their respective districts, leading me to conclude that findings likely generalize to these populations beyond my identification sample. I argue that results can inform recruitment and development efforts aimed at improving the quality of the teacher workforce.

The remainder of this paper is organized as follows. In the second section, I discuss previous research on the relationship between observational measures of teacher quality and student achievement. In the third section, I describe the research design, including the sample and data. In the fourth section, I present my identification strategy and tests of assumptions. In the fifth section, I provide main results and threats to internal and external validity. I conclude by discussing the implications of my findings for ongoing research and policy on teacher and teaching quality.

2. Background and context

Although improving the quality of the teacher workforce is seen as an economic imperative (Hanushek, 2009), longstanding traditions that reward education and training or offer financial incentives based on student achievement have been met with limited success (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Fryer, 2013; Harris & Sass, 2011; Springer et al., 2010). One reason for this posed by Murnane and Cohen (1986) almost three decades ago is the "nature of teachers' work" (p. 3). They argued that the "imprecise nature of the activity" makes it difficult to describe *why* some teachers are good and what other teachers can do to improve (p. 7).

Recent investigations have sought to test this theory by comparing subjective and objective (i.e., value-added) measures of teacher performance. In one such study, Jacob and Lefgren (2008) found that principals were able to distinguish between teachers in the tails of the achievement distribution but not in the middle. Correlations between principal ratings of teacher effectiveness and value added were weak to moderate: 0.25 and 0.18 in math and reading, respectively (0.32 and 0.29 when adjusted for measurement error). Further, while subjective ratings were a statistically significant predictor of future student achievement, they performed worse than objective measures. Including both in the same regression model, estimates for principal ratings were 0.08 standard deviations (sd) in math and 0.05 sd in reading; comparatively, estimates for value-added scores were 0.18 sd in math and 0.10 sd in reading. This evidence led the authors to conclude that "good teaching is, at least to some extent, observable by those close to the education process even though it may not be easily captured in those variables commonly available to the econometrician" (p. 103).

Two other studies found similar results. Using data from New York City, Rockoff, et al. (2012) estimated correlations of roughly 0.21 between principal evaluations of teacher effectiveness and value-added scores averaged across math and reading. These relationships corresponded to effect sizes of 0.07 sd in math and 0.08 sd in reading when predicting future student achievement. Extending this work to mentor evaluations of teacher effectiveness, Rockoff and Speroni (2010) found smaller relationships to future student achievement in math between 0.02 sd and 0.05 sd. Together, these studies suggest that principals and other outside observers understand some but not all of the production function that converts classroom teaching and professional expertise into student outcomes.

In more recent years, there has been a growing interest amongst educators and economists alike in exploring teaching practice more directly. This now is possible through the use of observation instruments that quantitatively capture the nature and quality of teachers' instruction. In one of the first econometric analyses of this kind, Kane, Taylor, Tyler, and Wooten (2011) examined teaching quality scores captured on the Framework for Teaching instrument as a predictor of math and reading test scores. Data came from Cincinnati and widespread use of this instrument in a peer evaluation system. Relationships to student achievement of 0.11 sd in math and 0.14 sd in reading provided suggestive evidence of the importance of general classroom practices captured on this instrument (e.g., classroom climate, organization, routines) in explaining teacher productivity.

At the same time, this work highlighted a central challenge associated with looking at relationships between