Economics of Education Review 40 (2014) 183–204

Contents lists available at ScienceDirect

Economics of Education Review

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

Skills, productivity and the evaluation of teacher performance☆

Douglas N. Harris [a], Tim R. Sass [b,*]

[a] Department of Economics, Tulane University, United States
[b] Department of Economics, Georgia State University, United States

ARTICLE INFO

Article history:
Received 6 June 2013
Received in revised form 6 March 2014
Accepted 10 March 2014
Available online 26 March 2014

JEL classification:
I21
J24

Keywords:
Teacher evaluations
Observational ratings
Value-added

ABSTRACT

We examine the relationships between observational ratings of teacher performance, principals' evaluations of teachers' cognitive and non-cognitive skills and test-score based measures of teachers' productivity. We find that principals can distinguish between high and low performing teachers, but the overall correlation between principal ratings of teachers and teachers' value-added contribution to student achievement is modest. The variation across metrics occurs in part because they are capturing different traits. While past teacher value-added predicts future value-added, principals' subjective ratings can provide additional information, particularly when prior value-added measures are based on a single year of teacher performance.

1. Introduction

Research consistently finds that teacher productivity is the most important component of a school's effect on student learning and that there is considerable heterogeneity in teacher productivity within and across schools.[1] The paramount role of teachers has led policymakers to focus on personnel policies governing selection, retention and compensation of teachers as a mechanism for enhancing educational quality.

At the heart of all teacher personnel policy decisions is the issue of how to evaluate teacher performance. Traditionally, teacher hiring, retention and salary decisions have been based on teacher credentials such as certification status, educational attainment and experience. However, except for the first few years of experience, research has failed to find a strong and consistent link between these measures and student outcomes.[2] Spurred on by the federal *Teacher Incentive Fund* (TIF) and *Race to the Top* (RTTT) initiatives many states and districts are

☆ This study is funded under grant R305M040121 from the U.S. Department of Education. We wish to thank Stacey Rutledge, William Ingle, Peter Hinrichs and participants in the NBER summer education workshop for their comments. We are also grateful to Brian Jacob for sharing his computer code to determine conditional probabilities of teacher performance. Abhir Kulkarni, Julia Manzella, John Gibson, William Ingle, Micah Sanders, Cynthia Thompson and Lisa Verdon provided valuable research assistance. Previous versions of the paper circulated under the title "What Makes for a Good Teacher and Who Can Tell?" All remaining errors are the responsibility of the authors.

* Corresponding author. Tel.: +1 404 413 0150.
E-mail addresses: dharri5@tulane.edu (D.N. Harris), tsass@gsu.edu (T.R. Sass).

[1] See, for example, Rockoff (2004), Hanushek, Kain, O'Brien, and Rivkin (2005), Rivkin, Hanushek, and Kain (2005), Kane, Rockoff, and Staiger (2008), and Aaronson, Barrow, and Sander (2007).

[2] See Rockoff (2004), Hanushek et al. (2005), Jepsen (2005), Rivkin et al. (2005), Boyd, Grossman, Lankford, Loeb, and Wyckoff (2006), Clotfelter, Ladd and Vigdor (2006, 2007, 2010), Kane et al. (2008) and Harris and Sass (2011). Harris and Sass (2011) find returns to experience beyond the first few years, particularly in middle school, and Wiswall (2013) uncovers high returns to later career experience at the elementary level.

http://dx.doi.org/10.1016/j.econedurev.2014.03.002
0272-7757/

beginning to incorporate both observations of teacher behavior and measures of student achievement in teacher evaluations.[3] They are also de-emphasizing, or in some cases eliminating, the use of traditional measures like attainment of a master's degrees and seniority in retention and compensation systems.

Despite the recent policy shift, little is known about the relative merits of observational measures and "value-added" ratings of teachers based on student test scores. Recent work by Chetty, Friedman, and Rockoff (2011) finds that students taught by high value-added teachers are more likely to have desirable long-run outcomes, including greater educational attainment, higher earnings and a reduced probability of teenage pregnancy. While no extant research links observational measures of teacher performance to student long-term outcomes, there is mounting evidence that observational measures of teacher quality are not strongly correlated with teacher value-added (Jacob and Lefgren, 2008; Mihaly, McCaffrey, Staiger, & Lockwood, 2013; Rockoff, Staiger, Kane, & Taylor, 2010, 2012).[4] Thus observational measures are not simply duplicative of value-added metrics and the divergence between the two suggests that observational measures could be capturing a different set of teacher skills, which could influence long-term student outcomes in ways that are not captured by value-added.[5]

Intertwined with the issue of how best to evaluate teacher performance is the relationship between teacher skills and teacher productivity. Recent work in labor economics suggests that both cognitive ability and non-cognitive personality traits, such as conscientiousness, play an important role in determining worker productivity (Borghans, ter Weel, & Weinberg, 2008b; Cunha, Heckman, Lochner, & Masterov, 2006; Heckman, Stixrud, & Urzua, 2006). Borghans et al. (2008b) theorize that different types of jobs require different combinations of personality traits and provide evidence that some of these traits are correlated with productivity. They find that "caring" is more important in teaching than in any other occupation, except nursing.

Personality traits are difficult to measure objectively (Borghans, Duckworth, Heckman, & ter Weel, 2008) and perhaps are more easily captured through direct observation. Thus if teacher observational rubrics measure non-cognitive traits that are not captured by teachers' contributions to student test scores, observational measures of teacher performance could serve as a valuable complement to value-added when evaluating teachers.

If observational measures are to be used in teacher evaluation systems there is also the issue of who can provide the most cost-effective evaluation of teacher performance. While some districts, like Washington DC, are utilizing trained observers to evaluate teachers, the vast majority of teacher evaluation systems being implemented across the country rely on the observations of principals to assess teacher performance. Although they may lack specific training in observational evaluation, principals may be lower-cost evaluators. Principals are typically required to observe teachers as part of their job and they collect a lot of information informally, and inexpensively, in the natural course of being in the school, interacting with teachers and talking to parents. Principals may also define performance somewhat differently to include contributions to output made through group interaction, e.g., mentoring of other teachers (Harris & Sass, 2011). Despite the widespread use and possible cost advantages of using principals to conduct observational evaluations of teacher, there is currently little evidence on whether some principals are better than others at evaluating teachers and whether the ability to evaluate teacher performance varies across different types of teachers or school environments.

In this paper we seek to enhance understanding of the relative merits of observational and value-added measures of teacher performance and shed light on the role that cognitive and non-cognitive skills play in determining teacher productivity. Specifically, we employ data on principals' evaluations of their teachers to address the following five questions:

1) How well do principal evaluations correlate with value-added measures of teacher productivity?
2) How does the ability of principals to measure teacher performance vary with the characteristics of principals and teachers?
3) What teacher traits are associated with their ability to promote student achievement?
4) Beyond the ability to raise achievement in the short-run, what traits do principals consider when evaluating teachers?
5) How well do principal evaluations and prior measures of teacher value-added predict *future* teacher productivity?

In the next section we describe the small existing literature on subjective evaluations of teachers and their relationship with value-added. This is followed by a discussion of the data used for our analysis, including how the interviews with principals were conducted and our method for estimating teacher value-added. In the concluding section we discuss our empirical results and possible policy implications.

2. Literature review

There is a limited literature that specifically addresses the relationship between subjective and objective

[3] A summary of the evaluation systems proposed by RTTT grantees is provided in Appendix Table A.1.

[4] The relationship between subjective ratings and objective performance measures is also relatively weak in other occupations (Bommer et al., 1995; Heneman, 1986).

[5] It is also possible that the two metrics measure the same underlying traits, but diverge because one or both are biased. Indeed, value-added measures are frequently criticized for potential selection bias due to non-random assignment of students to teachers (Rothstein, 2010). However, observational ratings of teachers could be subject to the same sort of bias if unobserved student characteristics affect the perceived performance of teachers. For example, if students with behavioral problems are more likely to be assigned to inexperienced teachers, raters could incorrectly perceive that less experienced teachers have poorer classroom management skills. In addition, observational ratings of workers could be subject to biases of observers (Varma & Stroh, 2001).