Economics of Education Review 36 (2013) 216–228

Contents lists available at SciVerse ScienceDirect

Economics of Education Review

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

# Is a good elementary teacher always good? Assessing teacher performance estimates across subjects

Dan Goldhaber *, James Cowan, Joe Walch

*Center for Education Data and Research, University of Washington – Bothell, 3876 Bridge Way North, Suite 201, Seattle, WA 98103, United States*

ARTICLE INFO

*Article history:*
Received 20 September 2012
Received in revised form 18 June 2013
Accepted 25 June 2013

*JEL classification:*
I20
I21

*Keywords:*
Educational economics
Productivity
Value-added
Teacher effectiveness
Human capital

ABSTRACT

In most elementary schools, teachers are responsible for several subjects. Various personnel policies, such as evaluating teachers based on value-added estimates aggregated across subjects or departmentalizing teachers, implicitly make assumptions about how closely teacher effectiveness is aligned across subjects. This paper reports on research exploring these issues using student–teacher linked data from North Carolina to assess the correlation of teacher productivity across math and reading. We find correlations of value-added estimates of about 0.6 and correlations in the underlying teacher effectiveness of 0.7–0.8. Assigning teachers to teach particular subjects based on their measured productivity could yield modest student achievement benefits.



## 1. The use of value-added teacher effect estimates

Policymakers are now using student growth-based measures of teacher effectiveness for a number of high-stakes personnel decisions. This policy direction is supported by research showing that teacher effectiveness varies widely and the variation has educationally meaningful consequences for student test achievement (Aaronson, Barrow, & Sander, 2007; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005).[1] Some research cautions about the use of value added, raising issues as to the validity and stability of effectiveness measures, as well as the possibility that teacher value-added effects "fade out" over time (Jacob, Lefgren, & Sims, 2010; Konstantopoulos, 2007; McCaffrey, Sass, Lockwood, & Mihaly, 2009; Rothstein, 2010).[2] But, recent research provides a measure of external validity to value-added estimates, showing that value-added estimates of individual elementary and middle school teachers are statistically significant predictors of college attendance and labor market earnings (Chetty, Friedman, & Rockoff, 2011).

Regardless of the academic debate about value-added, it seems clear that policymakers are likely to accelerate the use of student growth based measures to inform

* Corresponding author. Tel.: +1 206 547 5585; fax: +1 206 547 1641.
*E-mail address:* dgoldhab@u.washington.edu (D. Goldhaber).

[1] The literature typically finds teacher effect size estimates in the neighborhood of 0.10–0.25 standard deviations. The estimates are typically in the neighborhood of 0.10–0.15 for within-school estimates and are 0.15–0.25 for estimates that include between-school differences in teacher effectiveness. See, for instance, Goldhaber and Hansen (2012) and Hanushek and Rivkin (2010) for a more thorough discussion of the teacher effect size literature.

[2] There is an active debate over how to interpret findings on validity and stability and whether value-added measures ought to be used for personnel decisions. See, for instance, Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012), Glazerman et al. (2010), Goldhaber and Chaplin (2012), Hill (2009), and Harris (2009).

0272-7757/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.econedurev.2013.06.010

high-stakes personnel decisions such as tenure and compensation. Indeed, current policy initiatives such as Race to the Top and the Teacher Incentive Fund have created financial incentives for states and districts to incorporate these measures into their teacher evaluation systems. In light of their increasing use to classify teachers, there is a burgeoning literature that explores the correlation of these value-added estimates across time (Bill & Melinda Gates Foundation, 2011; Goldhaber, Gabele, & Walch, 2012; McCaffrey et al., 2009); across model specification (Ballou, Sanders, & Wright, 2004; Goldhaber et al., 2012; Kane & Staiger, 2008; Papay, 2011); across high and low stakes tests in the same subject area (Bill & Melinda Gates Foundation, 2011; Lockwood et al., 2007; Papay, 2011); and with other forms of teacher evaluation (Jacob & Lefgren, 2008; Kane, Taylor, Tyler, & Wooten, 2011). This literature generally shows teacher effect estimates are highly correlated across model specification. For instance, Goldhaber et al. (2012) find correlations of 0.4–0.95 across models that include different combinations of student covariates, student fixed effects, and school fixed effects. The adjacent year correlations are far lower, however, as previous studies have found correlations in the range of 0.2–0.5 with math value-added tending to be more stable than reading (Bill & Melinda Gates Foundation, 2011; Goldhaber et al., 2012). Finally, research using value-added estimated from different kinds of tests has found that estimates of teacher effectiveness are sensitive to changes in the testing instrument. An analysis from the Measures of Effective Teaching project estimates a correlation of 0.38 in value-added measures across mathematics tests and 0.22 across reading tests taken in the same year and subject (Bill & Melinda Gates Foundation, 2011). Taken as a whole, the literature suggests that teacher effectiveness exhibits educationally meaningful variation across test content and classrooms.

Few studies, however, look directly at the question of whether value-added measures are correlated across subjects. The correlations of math and reading value-added estimates for elementary and middle school teachers found in the literature range from 0.35 to 0.65 (Koedel & Betts, 2007; Loeb, Kalgorides, & Beteille, 2012; Teh, Resch, Walsh, Isenberg, & Hock, 2013; Value-Added Research Center, 2010). Value-added estimates are, of course, measured with error and correlation coefficients may not accurately represent the relationship in underlying effectiveness of teachers across subjects. Correcting for the regression error of teacher value-added estimates, Koedel and Betts (2007) find that estimated correlations increase from 0.35 to 0.64. Lefgren and Sims (2012) also implicitly find a positive relationship between math and reading effectiveness, although they do not directly estimate the correlation of teacher effectiveness across subjects. Specifically, they find there is a stable component to teacher effectiveness across years and subjects and that the ability of past value-added measures to predict teachers' future value-added increases when composite math and reading measures of value-added are utilized.

The gap in the literature on the underlying effectiveness of teachers across subjects is surprising. Under many evaluation systems for elementary school teachers, a key assumption is that teachers who are effective in one subject area also tend to be effective in other areas.[3] For instance, clearly it is important to know whether policies that reward or sanction teachers based on value-added are likely to be rewarding or sanctioning teachers who are effective or ineffective across the subjects they are responsible for teaching. In addition, school systems are increasingly using measures of teacher effectiveness for purposes other than high-stakes decisions.[4] One such possibility is to identify teachers as particularly effective at teaching certain subjects and departmentalize elementary schools, as is common at the middle- and high-school levels.[5] However, the potential gains from such a system clearly depend how much teacher effectiveness varies across subjects and how accurately these differences can be predicted.

In this paper we add to the sparse literature exploring teacher effectiveness across subjects. In particular, we use a 7-year panel of statewide data from elementary schools in North Carolina to estimate teacher value-added in math and reading using a variety of model specifications. We estimate correlations in estimated value-added measures within years across subjects of about 0.6, which is consistent with prior studies. We additionally use the estimated value-added in other years to correct for sampling error and other sources of non-persistent variation in teacher effectiveness, such as correlated random errors across tests within the same classroom, and find that the correlation in the persistent component of teacher value-added across subjects is approximately 0.7–0.8. We further find that these results are robust to changes in observable classroom characteristics and are reflected in average changes in student achievement when teachers move across schools and grades. Finally, we use the measured correlation between math and teaching competency to analyze the potential to use value-added assessments to assign teachers to subjects based on their measured effectiveness.

## 2. Data

We use administrative records collected by the North Carolina Department of Public Instruction (NCDPI) and managed by Duke University's North Carolina Education Research Data Center (NCERDC). These data include information on student performance on standardized tests in math and reading that are administered as part of the North Carolina accountability system. We standardize student test scores within grades and years. The student data also include individual information about students, such as gender, race and ethnicity, disabilities, and FRL status. In order to use a stable set of exams and covariates

[3] The terms "teacher value-added", "teacher effectiveness", and "teacher performance" are used interchangeably here.

[4] Examples include the use of value-added to assess the effectiveness of teacher training programs (Gansle, Noell, & Burns, 2012) and for eligibility for programs matching effective teachers to high-needs schools (Glazerman, Protik, Teh, Bruch, & Seftor, 2012).

[5] Recent proposals for departmentalizing elementary teachers include Hess (2009) and Public Impact (2012).