Teacher Effectiveness and the Achievement of Washington Students in Mathematics

CEDR Working Paper 2010-6.0

Dan Goldhaber
Center for Education Data & Research
University of Washington

Stephanie Liddle
Center for Education Data & Research
University of Washington

Roddy Theobald
Center for Education Data & Research
University of Washington

Joe Walch
Center for Education Data & Research
University of Washington

The research presented here utilizes confidential data from Washington State supplied by the Office of Superintendent for Public Instruction (OSPI). We gratefully acknowledge the receipt of these data. We also wish to thank Jordan Chamberlain for editorial assistance.

The suggested citation for this working paper is:

Goldhaber, Dan, Stephanie Liddle, Roddy Theobald, and Joe Walch. *Teacher Effectiveness and the Achievement of Washington Students in Mathematics.* Working paper no. 2010-6.0. Center for Education Data & Research (CEDR), University of Washington, 2010. Web. <http://www.cedr.us/publications.html>.

I. A Focus on Teacher Quality

Research has clearly shown that a quality teacher can have a tremendous impact on student learning. For example, some studies show that the difference for students between having a very effective or very ineffective teacher can represent more than a year's worth of learning growth. But evidence on which specific teacher characteristics predict classroom effectiveness remains, to a large extent, an empirical mystery. This is a problem for policymakers who wish to improve the quality of the teacher workforce, since there aren't clear actions (e.g., licensure or pay policies) that would lead to improvements in the workforce.

The recent finding by a number of newer research studies that there appears to be a great deal of variation in the effectiveness levels of teachers in the workforce further complicates matters.[1] This implies that the distribution of teachers across students can have profound effects on educational achievement gaps.

Stories about the importance of teacher effectiveness for student achievement are currently ubiquitous in the media (Gladwell 2008; Kristoff 2009; Felch et al 2010). Teacher quality is also the focus of significant national efforts and investments.[2] Federal programs like Race to the Top and the Teacher Incentive Fund represent ambitious attempts to recognize, reward, and encourage effective teaching.[3] The federal investment, totaling approximately $100 billion,[4] is mirrored by similar philanthropic efforts. The Bill and Melinda Gates Foundation, for instance, invested hundreds of millions of dollars in an attempt to jump start teacher policy reforms in a number of "deep dive" school districts (districts in which Gates is working closely to implement changes to teacher policy and to assess the implications of those changes), and its Measures of Effective Teaching study focused on assessing the relationship between various methods of evaluating teachers and student achievement.[5]

A growing body of quantitative research supports the focus on teacher and teacher effectiveness. This research shows teaching to be the most important school-based factor influencing student performance (Aaronson et al., 2007; Rivkin et al., 2005). The means to improve the effectiveness of the teacher workforce, however, is not straightforward; experience,

---

[1] See, for instance, Aaronson et al. (2007), Gordon et al. (2006), and Rivkin et al. (2005).
[2] We use the terms "teacher effectiveness" and "teacher quality" interchangeably here to mean the *ability of teachers to contribute in measurable ways to student gains on standardized tests*. While helping students to learn so that they perform better on tests is a narrow conception of the totality of what teachers do, it is useful to define teacher quality in this way because it provides a means of measuring teachers based on student outcomes. As we go on to describe below, other input-based definitions of teacher quality (e.g. a teacher's credentials) are not strongly related to student achievement. Moreover, student performance on tests has been shown to be an important predictor of later life outcomes (Grogger and Eide, 1995; Murnane et al., 1995), is related to aggregate country growth rates (Hanushek, 2009), and in many countries, helps to drive educational policies (e.g. school accountability).
[3] The Race to the Top grant program specifically encourages states to use teacher performance to inform "key decisions in such areas as evaluation and development, compensation and advancement, tenure and removal" (Weiss 2009), while the primary goal of the Teacher Incentive Fund is "improving student achievement by increasing teacher and principal effectiveness."
See: http://www2.ed.gov/programs/teacherincentive/index.html
[4] The American Recovery and Reinvestment Act (2009) provided the majority of the funding for these federal programs. See http://www2.ed.gov/policy/gen/leg/recovery/implementation.html.
[5] For more information on the deep dive effort, see Phillips (2008),) and for more on the Measures of Effective Teaching study, see Bill and Melinda Gates Foundation (2010).

degrees, and credentials—factors that typically determine teacher employment eligibility and compensation—do not adequately explain effectiveness. In addition, teacher in-service evaluation systems typically do not recognize the significant variation we know exists amongst teachers (Goldhaber, 2010).

In this working paper we summarize findings on the importance of teacher effectiveness for students' elementary level mathematics achievement in Washington State. Washington, like many states, has recently implemented reforms designed to identify and enhance the quality of its teachers. For instance, Senate Bill 6696 requires principals to use a four-level evaluation system rather than a binary rating system and to evaluate teachers for at least 90 minutes during the school year; it also extends the pre-probationary period for new teachers from two years to three years. Understanding the efficacy of efforts like these is essential: consistent with findings elsewhere in the country, the impact of teacher quality on Washington's students is quite significant.

II. Data and Methodology

The findings on the effects of teachers on elementary level student achievement in math are based on a unique dataset from Washington State linking teachers to their schools and students. Due to the lack of a centralized data collection repository, linking data on Washington State teachers and students necessitates that we coordinate a number of variables across a variety of data sets. The Center for Education Data and Research (CEDR) stores and maintains a unique collection of Washington State education data for the purposes of educational policy research. Reconciled from a number of state administrative longitudinal databases, this data repository contains historical information about schools, teachers, and students.

Using this repository, we have currently matched Washington State public school students in grades 3–6 to teachers across four school years (2005–06 through 2008–09).[6] To date, across all years and grades, we have matched 10,901 teachers (23,011 teacher-years) to 170,981 students (403,445 student-years).

These data allow us to estimate value-added models (VAMs) of teacher effectiveness based on students' achievement on the Washington State Assessment of Student Learning (WASL)—a regular annual statewide assessment test for grades 3rd–8th grade and 10th grade. These tests are used to assess student performance, determine promotion, and evaluate schools and teachers in the state. The sample we use for this study includes all teachers for whom we can calculate VAM estimates, who have at least 10 students with valid test scores, leaving a sample of 6,731 teachers (12,134 teacher-years) and 163,938 students (222,006 student-years).

[6] The proctor of the state assessment was used as the teacher-student link for at least some of the data used for analysis. The 'proctor' variable was not intended to be a link between students and their classroom teachers so this link may not accurately identify those classroom teachers. It is possible to link student-level WASL files to teacher-level S275 files. We do this by matching the proctors in the WASL files to teachers in the S275 files using the district code, school code, and last names. That is, the proctor must be located in the same district and school as the teacher and have the same last name. This match is made primarily made within the same year across the two types of files. However, to account for the inconsistencies and variations in names across yearly waves of the S275 data, if a match cannot be found in the same year, the proctor names are matched to the previous year and following year of S275. The WASL proctors names are then updated with correct spellings and matched to teachers in the S275 data. The table below shows the result of this match.

There is a growing body of literature that uses VAMs in an attempt to identify the contribution that *individual* teachers make toward student learning gains (e.g. Aaronson et al., 2007; Rivkin et al., 2005). The basic idea behind using value-added is to predict the achievement levels of students based on: 1) their prior achievement; 2) their background characteristics (e.g. whether they are receiving free or reduced price lunch); and 3) schooling resources (e.g. per pupil expenditure and class size). Differences between the prediction of student achievement and actual achievement are attributed to teachers and these differences, aggregated to the classroom level, are treated as a measure of teachers' performance or effectiveness.

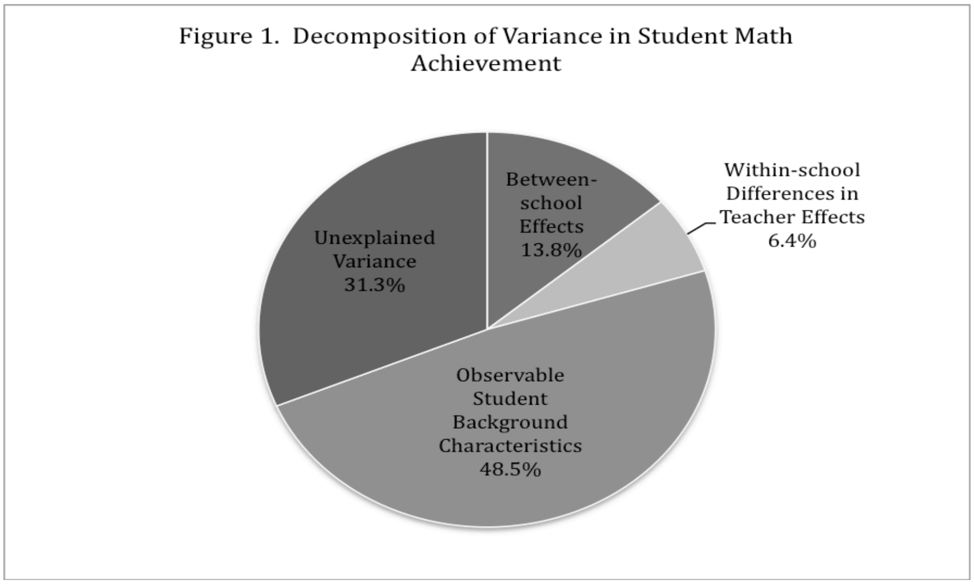
There is no universally accepted method for calculating a teacher's value-added contribution and research shows that methodology and context can influence the measure (Ballou et al. 2004, Goldhaber and Hansen, 2010; McCaffrey et al., 2004; Rothstein, 2010; Rubin, Stuart, and Zanutto, 2004; Tekwe et al., 2004). The results we report here are based on a measure that uses as many years of student-teacher data that are available to inform each individual teachers' effectiveness estimate, and it shrinks these estimates toward the overall mean effect in proportion to its reliability.[7] However, while below we report teacher effects based on a particular method of calculating them, the results are not exceedingly sensitive to the particular (value-added model) way in which they are estimated (Goldhaber and Theobald, 2010).

III. Findings

Our student achievement model explains about 70 percent of the variation in students' mathematics achievement on the WASL tests; a figure that is fairly typical of the literature (Jackson and Bruegmann, 2009; Jacob, Lefgren, and Sims, 2008). And while teacher effectiveness is found to be an important factor in explaining student achievement, teacher characteristics and credentials do little to predict what makes teachers effective or ineffective; this too is consistent with the educational productivity literature (Hanushek, 1997).[8] We illustrate this in **Figure 1**, which shows the decomposition of the variance in student achievement into: 1) observable student background characteristics (including prior student achievement); 2) the within-school differences in teacher effects; 3) the between-school effects; and 4) the unexplained portion of the variance (or error).

[7] For more detail on this Empirical Bayes methodology, see Aaronson et al. (2007).

[8] At least in terms of value-added measures of effectiveness. Teachers may well be contributing to student learning in ways that are not captured by state assessments.



The above picture shows that what students bring to the table at the beginning of the year (their characteristics and prior level of achievement in a subject) is vitally important, but school quality matters a great deal too, explaining about 20 percent of the variation (the combination of school-level and teacher-level effects). Of this, over 6 percent is attributed to within school differences in teacher effectiveness. This is a very conservative estimate of the importance of differences between teachers because some of the between-school variation is also due to differences in teacher effectiveness; unfortunately we cannot separate differences between schools in the effectiveness of the teacher workforce from other school-level factors (e.g. principals) that might also influence achievement.[9]

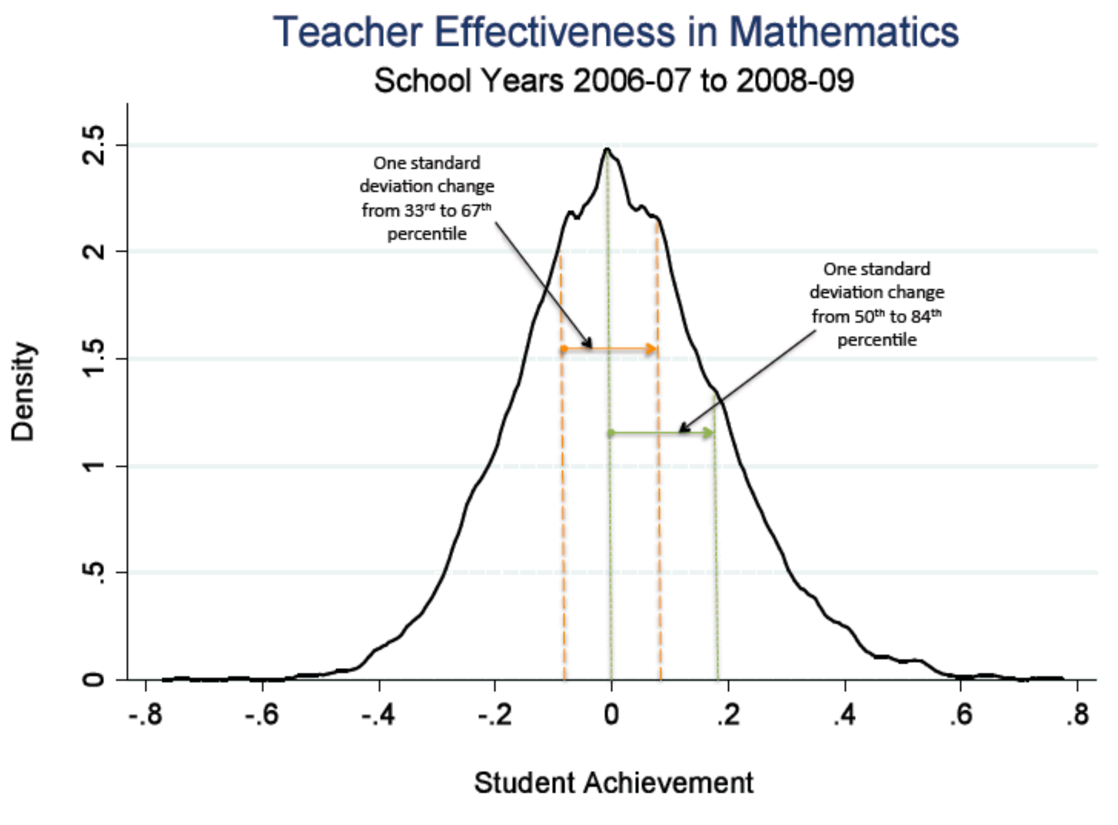
So what are the consequences for students of having an effective teacher versus an ineffective teacher? **Figure 2** illustrates the importance of teacher effectiveness for student achievement showing the estimated distribution of teacher effectiveness projected on student achievement.[10] Our findings suggest that a one standard deviation increase in teacher effectiveness (for instance, the difference between a teacher at the 33rd percentile of the effectiveness distribution and one at the 67th percentile or the difference between a teacher at the 50th percentile of the effectiveness distribution and one at the 84th percentile of the distribution) would increase student achievement by about 18 percent of a standard deviation.[11]

[9] These estimates of the relative importance of various factors in explaining the variation in student achievement are extremely similar to estimates from elsewhere. Goldhaber et al. (1999), for instance, peg this figure at 8.5 percent; and Nye et al. report that differences between teachers explain between 10.4 and 11.3 percent of the variation in mathematics scores for students in kindergarten through third grade.

[10] These are kernel density distributions.

[11] These are estimates that are "shrunken" toward the overall mean of teacher effectiveness in proportion to the reliability of the estimates. The unadjusted estimate of a one standard deviation change in effectiveness is about 22 percent of a standard deviation in student achievement. Note, however, that because the distribution of teacher effectiveness is not perfectly normal and is slightly skewed to the right, not every one standard deviation change in teacher effectiveness results in the same effect on student achievement. In general, changes in teacher effectiveness on the left, or negative, side of the distribution result in smaller changes in student achievement compared to the right, or positive, side of the distribution. For example, moving from the 0.15th to the 2.5th percentile of teacher

Figure 2.



To put this in perspective, students gain nearly a full standard deviation in math achievement as they advance from one grade to the next in upper elementary grades, so the difference we estimate is equivalent to roughly 2.6 months of learning (Schochet and Chiang, 2010). Alternatively, the typical gap between black and white students or economically advantaged and disadvantaged students is in the range of 0.7 to 1 standard deviations (Hanushek and Rivkin, 2010) so our findings suggest that having a highly effective teacher (i.e., in the 84th percentile) rather than an average teacher, could cut these achievement gaps down by nearly one fifth of a standard deviation.

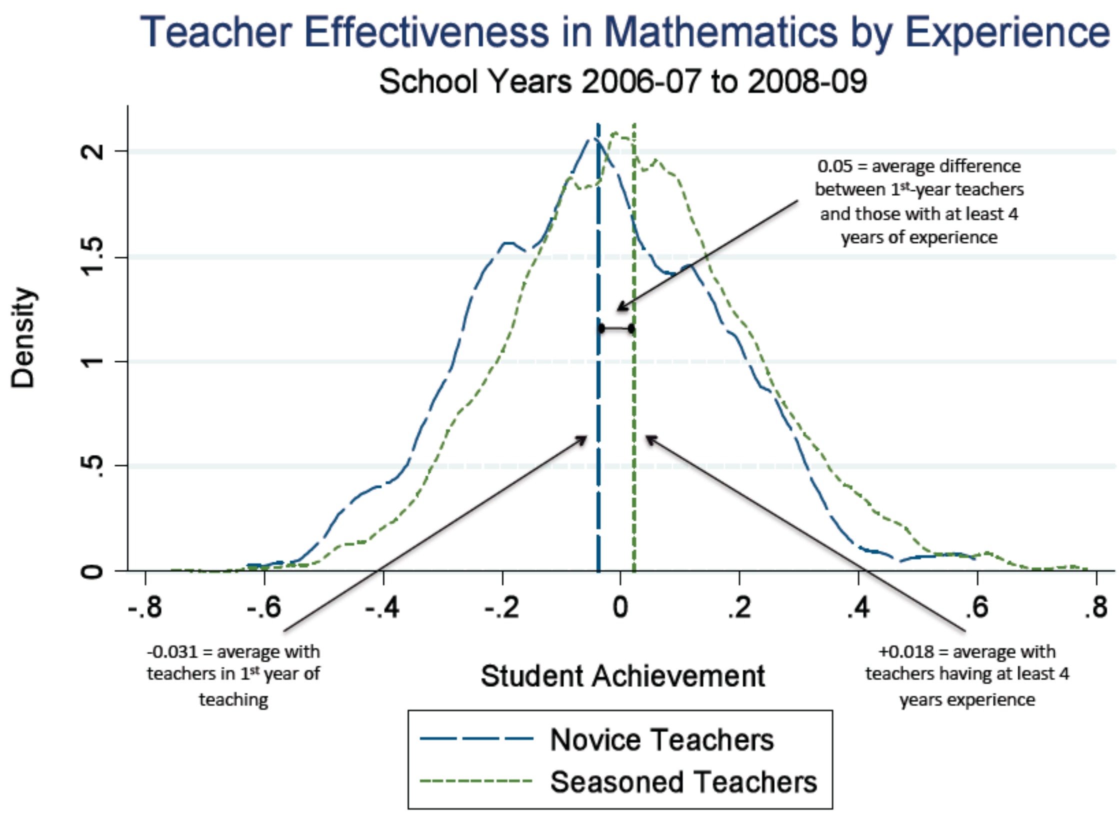
We attempt to explain the differences between teachers as a function of teacher characteristics including: sex, race/ethnicity, education, teaching experience, and training institution. While many of these variables are statistically significant (though not always in the expected direction), collectively they account for only about 2 percent of the variation in teacher effectiveness.[12] Teacher experience stands out as the most important factor in predicting teacher effectiveness, a common finding (Clotfelter et al., 2006; Goldhaber and Hansen, 2010; Rivken et al., 2004; Rockoff, 2004). But while experience is predictive, it is by no means strongly

effectiveness results in a 17.3% change in student achievement, moving from the 2.5th to the 16th percentile=15.3% change, from the 16th to the 50th percentile=15.5% change, from the 50th to the 84th percentile=17.1% change, from the 84th to the 97.5th percentile=19.3% change, and from the 97.5th to the 99.85th percentile=24.5% change in student achievement.

[12] These findings are strikingly similar to those based on a randomized experiment (Nye et al., 2004).

predictive when we look, for example, at the distributions of effectiveness for novice (1st year) teachers as compared to more seasoned teachers, i.e. those with 4 or more years of experience (**Figure 3**).

Figure 3.



Our models suggest that, all else equal, students with novice teachers would score about 3 percent of a standard deviation lower on the math WASL than students with teachers with average experience. And, by contrast, students assigned to teachers with four or more years of experience would score about 2 percent of a standard deviation higher than those with a teacher with average experience. This 5 percent of a standard deviation difference between novice and seasoned teachers is certainly significant, but it pales in comparison to the differences that exist between teachers with the same experience level. For example, the difference between teachers that are a half of a standard deviation away from the mean in either experience category is 3.6 times the average difference between these experience groups. The bottom line is that while experience matters, there are certainly lots of novice teachers who are more effective than the average seasoned teacher and vice versa.

IV. Policy Implications

The findings in this working paper are not terribly novel, but they aptly illustrate that findings from a variety of other states and localities also apply in Washington State. Differences in teacher effectiveness have consequential implications for student achievement, but they are not well explained by teacher credentials. This means that current teacher policies—from licensure to compensation—which are built almost entirely around these credentials, are largely disconnected from teacher productivity. Unfortunately, research suggests that in-service evaluations of teachers typically also fail to recognize the meaningful differences that exist between teachers (Goldhaber, 2010; Weisburg et al., 2009).

For many, findings like those reported here imply the necessity of a significant national investment in new ways to assess teacher effectiveness and connect it to teacher policy. Ultimately, however, we will only be able to learn the implications of new teacher effectiveness policies through examination of actual policy variation. This creates a bit of catch-22 since the implementation of new policies is somewhat risky, both in terms of knowning the effects of a new policy and politically.[13] For instance, there is growing interest in using value-added methodologies to inform high-stakes teacher policies (Goldhaber, 2010), which is controversial (Baker et al., 2010). Value-added is certainly not a perfect system for evaluating teachers and it could lead some teachers to behave in a dysfunctional way that does not increase student learning (e.g. try to manipulate test results to produce better scores, or not cooperate with colleagues). However, the risks that may arise with any changes in policy must be judged against the current state of teacher assessment: a system that allows for some students to be educated by teachers clearly known to be ineffective and largely fails to recognize teacher excellence.

[13] Several high-profile reports (Baker et al. (2010); Braun et al. (2010); Schochet and Chiang (2010)), for instance, caution against too heavy a reliance on value-added measures of teacher effectiveness.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. Teachers and student achievement in the Chicago Public High Schools. *Journal of Labor Economics* 25:95-135.

Ballou, Dale, William L. Sanders, and Paul S. Wright. 2004. "Controlling for Student Background in Value-Added Assessment of Teachers. *Journal of Education and Behavioral Statistics* 29:37-65.

Bill and Melinda Gates Foundation. 2010. "Working with Teachers to Develop Fair and Reliable Measures of Effective Teaching."

Baker, Eva, Paul Barton, Linda Darling-Hammond, Edwrad Haertel, Helen Ladd, Robert Linn, Diane Ravitch, Richard Rothstein, Richard Shavelson, and Lorrie Shepard. 2010. "Problems with the Use of Student Test Scores to Evaluate Teachers." *Economic Policy Institute*.

Braun, Henry, Naomi Chudowsky, and Judith Koenig (Eds). 2010. "Getting Value out of Value-Added: Report of a Workshop." Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Accountability; National Research Council.

Clotfelter, Charles, Lennie, Ladd, and Vigdor. 2006. "Would Higher Salaries Keep Teachers in Schools? Evidence from a Policy Intervention in North Carolina." *National Bureau of Economic Research.* Working Paper. June 2006.

Felch, J., Song, J., and Smith, D. "Who's Teaching L.A.'s Kids?" *The Los Angeles Times,* August 14, 2010.

Gladwell, Malcolm. 2008. "Most Likely to Succeed." *The New Yorker*. December 15, 2008.

Goldhaber, Dan. 2010. When the Stakes are High, Can We Rely on Value-Added?: Exploring the Use of Value-Added Models to Inform Teacher Workforce Decisions. *Center for American Progress.*

Goldhaber, Dan, Dominic Brewer, and Deborah Anderson. 1999. "A Three-way Error Components Analyses of Educational Productivity." *Education Economics*. Vol. 7, No. 3, pp. 199-208.

Goldhaber, Dan and Michael Hansen (2010). "Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance." *Center for Education Data and Research*. Working Paper 2010-3.

Goldhaber, Dan and Roderick Theobald, 2010. "Assessing the Determinants and Implications of Teacher Layoffs." *Center for Education Data and Research.* Working Paper 2010-5.

Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger, 2006. *Identifying Effective Teachers Using Performance on the Job*. Washington, DC: *The Brookings Institute*.

Grogger, Jeff, and Eric Eide. 1995. "Changes in College Skills and the Rise in the College Wage Premium." *Journal of Human Resources* 30(2): 280-310.

Hanushek, Eric A. 1997. “Assessing the Effects of School Resources on Student Performance: An Update.” *Educational Evaluation and Policy Analysis* 19(2), 141-164.

Hanushek, Eric A. 2009. “Teacher Deselection.” Pgs. 165-180 in *Creating a New Teaching Profession*, Dan Goldhaber and Jane Hannaway, eds.

Hanushek, Eric A., Steven G. Rivkin 2010. “Generalizations About Using Value-Added Measure of Teacher Quality. *American Economic Review*. May, 2010.

Jackson, C. Kirabo, and Elias Bruegmann. 2009. “Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers.” *American Economic Journal: Applied Economics* 1:85-108.

Jacob, Brian A., and Lars Lefgren. 2008. “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics* 26:101-136.

Kristoff, Nicholas. 2009. “Our Greatest National Shame.” *The New York Times*. February 14, 2009.

McCaffrey, Daniel, J.R. Lockwood, Daniel M. Koertz, and Laura S. Hamilton. 2004. "Models for Value-Added Modeling of Teacher Effects", *Journal of Educational and Behavioral Statistics*, 29(1).

Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The growing importance of cognitive skills in wage determination." *Review of Economics and Statistics,* 77,no.2 (May):251-266.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. “How Large Are Teacher Effects?” *Educational Evaluation and Policy Analysis*. Vol. 26, No. 3, pp.237-257.

Phillips, Vicki. 2008. “A Forum on Education in America.” *Gates Foundation*.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. “Teachers, Schools, and Academic Achievement.” *Econometrica*. Vol. 73, No. 2, pp. 417-458.

Rockoff, J. E. 2004. “The Impact of Individual Teachers on Students' Achievement: Evidence from Panel Data.” *American Economic Review,* 94(2), 247-252.

Rothstein, Jesse. 2010. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *Quarterly Journal of Economics,* 125(1), 175-214.

Schochet and Chiang. 2010. “Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains.”

Tekwe, Carmen, Randy Carter, Chang-Xing Ma, James Algina, Maurice Lucas, Jeffery Roth, et al. 2004. "An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance." *Journal of Educational and Behavioral Statistics*, 29(1), 11-36.

Weisburg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling. 2009. "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness." Report for the *New Teacher Project*.

Weiss, Joanna. 2009. "Education's 'Race to the Top' begins." *Education Week*, July 23, 2009.