



BRIEF 10

MAY 2010

Using Performance on the Job to Inform Teacher Tenure Decisions

Dan Goldhaber and Michael Hansen

Well over a decade into the standards movement, the idea of holding schools accountable for results is being pushed to a logical, if controversial, end point: the implementation of policies aimed at holding individual teachers (not just schools) accountable for results. As a number of states begin to revamp their tenure-granting policies, the idea that high-stakes personnel decisions need to be linked to direct measures of teacher effectiveness (as a form of quality control in the workforce) is gaining traction among education policymakers.¹

The idea that high-stakes personnel decisions need to be linked to direct measures of teacher effectiveness is gaining traction among education policymakers.

The focus on teacher tenure reform is appropriate and timely. Race to the Top encourages states to adopt policies that measure the impact of individual teachers on student learning and use those measures to inform human capital decisions including tenure and compensation. Also, three important findings in teacher quality research underscore the need for reform: (1) Teacher quality (measured by estimated teacher impacts on student test score gains) is the most important school-based factor when it comes to improving student

achievement,² (2) Teacher quality is a highly variable commodity (Kane, Rockoff, and Staiger 2008); and (3) A strikingly small percentage of tenured teachers is ever dismissed for poor performance (Weisberg et al. 2009).

In recent months a number of states, such as Tennessee, have considered tying teacher evaluations and tenure to student achievement as part of their Race to the Top plans.³ This research brief evaluates how well early-career performance signals teacher effectiveness after tenure. The brief presents selected findings from a larger study using North Carolina data that examines the stability of value-added model (VAM) estimates and their value in predicting student achievement (Goldhaber and Hansen 2010). This research has important implications for policies relying on VAM estimates to control teacher quality in the workforce, given that a degree of stability of teacher performance over time is implicitly assumed.

ESTIMATING TEACHER PERFORMANCE AND ITS STABILITY: DATA AND ANALYTIC APPROACH

The administrative data we use is collected by the North Carolina Department of Public Instruction and includes information on all teachers and students in North Carolina in the school years 1995–96 through 2005–06. We restrict our analyses to students who are in self-contained classrooms (i.e., with the same teacher for the entire day) in grades 4 or 5 and who have valid end-of-year assessment scores in math and reading.⁴

Because VAM measures of teacher effectiveness are stable, early-career estimates predict student achievement at least three years later, and they do so far better than observable teacher characteristics.

Our findings rely on the estimation of teacher effectiveness. To obtain these estimates, we employ a commonly used VAM:

$$(1) A_{ijgt} = \alpha \mathbf{A}_{i(t-1)} + \mathbf{X}_{it}\gamma + \mathbf{T}_{\tau_j} + \mathbf{G}\delta_g + \mathbf{Y}\varphi_t + \varepsilon_{ijgt}$$

where math achievement of student i assigned to teacher j in grade g and year t is a linear function of prior achievement on both math and reading tests, $\mathbf{A}_{i(t-1)}$, a vector of student and family background characteristics, \mathbf{X}_{it} , and teacher, \mathbf{T} , grade, \mathbf{G} , and year, \mathbf{Y} , indicator variables. We use a rolling two-year window to estimate teacher effectiveness for all teachers at various points within the 11-year panel. The parameter estimates $\hat{\tau}_j$ provide a teacher-specific measure of effectiveness during each period.⁵

For the tenure analyses, we further restrict our analysis sample to teachers whose performance is observed both pre- and post-tenure. In North Carolina, state policy dictates that teachers receive tenure after teaching in the same district in the state's public schools for four consecutive years (Joyce 2000).⁶ In principle, using all four years of teacher job performance to grant tenure is possible, but in practice it is unlikely that four years of value-added calculations would be available for making a tenure decision.

Additionally, in many states, tenure is granted after just three years of classroom teaching (and in some states even sooner). For these reasons we focus on teacher effectiveness estimates based on a teacher's first two years of employment in a district, which will be used as an explanatory variable in predicting student performance in teachers' post-tenure period. The analysis sample includes 609 unique teachers and 26,280 teacher-student-year observations in the post-tenure period (most teachers are observed more than once in this period).

To assess the extent to which past teacher performance predicts student achievement, we estimate a variant of (1) for this sample of post-tenure teachers. In this model, we substitute a vector of teacher quality variables for the teacher indicator variables in (1), which includes $\hat{\tau}_j$, our estimated VAM measure of teacher effectiveness from *at least three years prior*, and we also include other school characteristics, such as class size.⁷

HOW WELL DO EARLY-CAREER TEACHER VAM ESTIMATES PREDICT STUDENT ACHIEVEMENT?

Table 1 shows the estimated coefficients for models that include either a set of observable teacher characteristics (column 1), VAM measures of teachers' early career performance (column 2), or both observable characteristics and VAM measures (column 3).

Consistent with a good deal of empirical literature (e.g., Clotfelter et al. 2007; Hanushek 1997), most teacher characteristics are not individually statistically significant; however, an F-test does indicate that they are jointly significant. The coefficient estimates on the pre-tenure teacher VAM estimate in column 2, by contrast, are highly significant.⁸ The point estimates suggest that a 1 standard deviation increase in a teacher's lagged effectiveness increases students' achievement scores by about 9 percent of a standard deviation.⁹ In column 3, we report on specifications that include both observed teacher variables and prior VAM estimates. In these models, the observable teacher variables are no longer jointly significant, and the estimates of the predictive power of lagged teacher effects are little changed.

These results suggest VAM teacher effect estimates are better indicators of teacher quality (at least as measured by standardized tests) than observable teacher attributes, even with a three-year lag between the time that the estimates are derived and student achievement is predicted.

Using VAM estimates to inform tenure decisions is not without costs, political or otherwise. For policy purposes, it is useful to understand the extent to which these estimates outperform other means of judging teachers. We explore this issue by comparing out-of-sample predictions of student achievement (based on models with observable teacher characteristics and predictions of

Table 1. Predicting Student Achievement Using Teacher Characteristics or Pre-tenure Job Performance Estimates

Teacher variables	(1)	(2)	(3)
6–9 years experience	0.016 (0.013)	0.005 (0.012)	
> 9 years experience	0.083 (0.048)	0.037 (0.050)	
Holds master's degree or higher	0.004 (0.021)	0.002 (0.018)	
Average licensure test score	0.015 (0.012)	0.016 (0.010)	
College selectivity	-0.002 (0.008)	-0.004 (0.007)	
Fully licensed	0.078 (0.051)	0.089 (0.053)	
Pre-tenure VAM teacher effects		0.091** (0.009)	0.091** (0.009)
R-squared	0.72	0.73	0.73

Notes: Standard errors in parentheses. All models include the following controls: a student's pre-test score in both math and reading, race/ethnicity, gender, free/reduced-price lunch (FRL) status, parental education, class size, percentage of minority students in the class, and percentage of students receiving FRL in the class. The omitted experience category is 5.

** Difference significant at the 1% confidence level.

achievement based on teacher effectiveness) to actual realized student achievement. We use the coefficient estimates from table 1 to predict student achievement in the school year 2006–07 for students who were enrolled in classes taught by teachers in the sample used to generate the results reported.¹⁰ For each student we obtain two different estimates

of achievement. The first is based on using teacher characteristics in the model (all those characteristics reported in column 1) and the second is based on the pre-tenure VAM measure of teacher effectiveness (column 2).

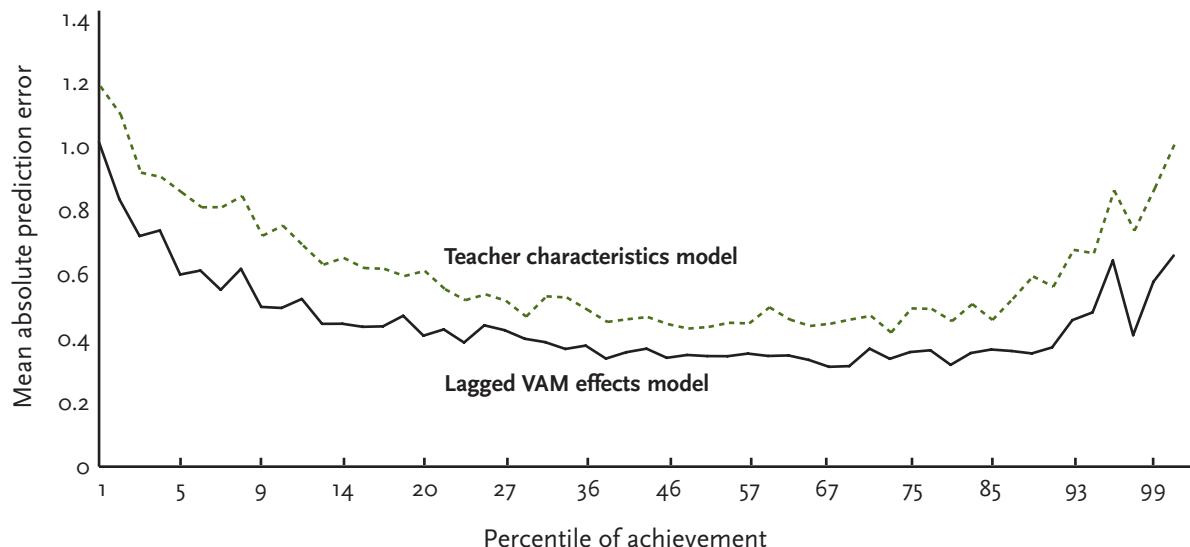
The pre-tenure VAM model has superior out-of-sample predictive power compared with the model that was based on teacher characteristics, as indicated by t-tests of the differences in mean absolute error between the observed student achievement and the predictions from the two models. To get a better sense of whether the differences between the VAM estimates and teacher observable estimates are meaningful, we plot the mean absolute error against actual student achievement.

Figure 1 shows the mean absolute error of predictions from both models for each percentile of achievement.¹¹ As might be expected, the results of this exercise show that both models do a relatively poor job of predicting student achievement far from the mean, but it also shows that the specification that includes pre-tenure VAM performance estimates is far superior to the specification that includes teacher characteristics variables.

POLICY IMPLICATIONS AND CONCLUSIONS

What would it mean to use VAM estimates to selectively “deselect” teachers before granting tenure (Gordon, Kane, and Staiger 2006; Hanushek 2009)? To provide some perspective,

Figure 1. Prediction Error as a Function of Student Achievement



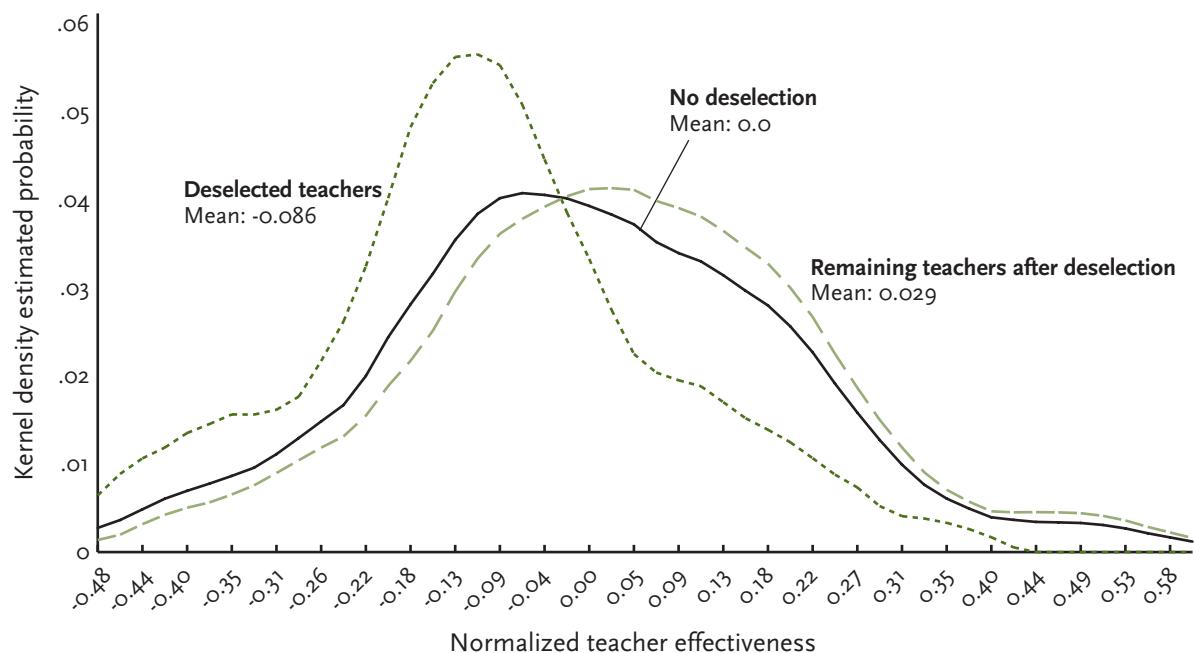
we calculate the effect on the post-tenure teacher workforce if the teachers with early career VAM effect estimates in the bottom quarter of the distribution were deselected.¹² Our calculations suggest that imposing this hypothetical rule would, *ceteris paribus*, have an educationally significant effect on the distribution of teacher workforce quality. This is illustrated in figure 2, which shows the teacher effectiveness distributions for the teachers in this sample based on their fifth year of teaching in the district.

The three distributions depicted are the estimated post-tenure effects for (1) deselected teachers, (2) the distribution with no deselection, and (3) the upper 75 percent of teachers retained after the filter is imposed. Deselected teachers are estimated to have mean impacts that are over 11 percent of a standard deviation of student achievement lower than retained teachers. The difference between the distribution of retained teachers and the distribution with no deselection is about 3 percent of a standard deviation of student achievement. When we take this a step further and replace deselected teachers with teachers who have effectiveness estimates equal to the average effectiveness of teachers in their first and second years, the post-tenure distribution average is still predicted to be 2.5 percent of a standard deviation higher than if teachers had not been deselected.

While these effects may appear small, new evidence suggests that even small impacts on teacher workforce quality can improve overall student performance and profoundly affect aggregate country growth rates. Economist Eric Hanushek (2009) estimates that a modest policy identifying and replacing 6–10 percent of the least effective teachers from the classroom could, over 20 years, improve the nation's gross domestic product by 1.6 percent—just about equal to the aggregate spending on current teacher salaries and benefits.

There has been immense debate over policies designed to enhance the quality of teachers. Recent evidence that observable teacher characteristics are only weakly related to teacher productivity makes current teacher quality policies elusive, leading some education policymakers and researchers to call for using more direct measures of teacher performance to determine employment eligibility (or compensation). The evidence presented here shows that VAM measures of teacher effectiveness are stable enough that early-career estimates of teacher effectiveness predict student achievement at least three years later, and that they do so far better than observable teacher characteristics. This finding reinforces the notion that these estimates are a reasonable metric to use as a factor in making substantive personnel decisions.

Figure 2. Teacher Quality Distribution of Tenure Decision Subgroups



Of course, there are reasons to proceed with caution. The limitations of our analyses are that they are based on a very restrictive sample, so the findings we present may not be generalizable to the teacher workforce. Moreover, our calculations are only based on a partial equilibrium analysis: using VAM to inform tenure decisions would represent a seismic shift in teacher policy. Such a shift could have far-reaching consequences for who opts to enter the teacher labor force and how teachers in the workforce behave.

Even small impacts on teacher workforce quality can improve overall student performance and profoundly affect aggregate country growth rates.

Today, teaching is a relatively low-risk occupation as salaries are generally determined by degree and experience levels, and job security within the field is high. Policies that make the occupation more risky might induce different types of entrants, but economic theory also suggests that teacher quality will only be maintained if salaries are increased enough to offset any increased risk associated with becoming a teacher. In sum, we cannot know the full impact of using VAM-based reforms without assessments of actual policy variation.

NOTES

1. Empirical studies have shown that licensure requirements raise the barriers to entry into the teaching profession but often do little to control the overall quality of the workforce (Goldhaber 2007; Kane, Rockoff, and Staiger 2008). Tenure attaches considerable employment protections to teachers, but anecdotal evidence suggests rewarding it to teachers is commonly more procedural than a rigorous quality check (Jason Felch, Jessica Garrison, and Jason Song, "Bar Set Low for Lifetime Job in L.A. Schools," *Los Angeles Times*, December 20, 2009, accessible at <http://www.latimes.com/news/local/education/la-me-teacher-tenure20-2009dec20,0,2529590.story>).
2. Rivkin, Hanushek, and Kain (2005) and Rockoff (2004) estimate that a 1 standard deviation increase in teacher quality raises student achievement in reading and math by about 10 percent of a standard deviation—an achievement effect on the same order of magnitude as lowering class size by 10 to 13 students (Rivkin et al. 2005).
3. See, for instance, "Bresden Will Give Legislators a Week to Pass Education Reform," *Tennessee Journal*, December 18, 2009, 1–2.
4. The North Carolina data do not include explicit ways to match students to their classroom teachers. They do, however, identify the proctor of each student's end-of-grade tests, and in elementary school the exam proctors are generally the teachers for the class. We use the listed proctor as our proxy for a student's classroom teacher but take several precautionary measures (described in greater detail in Goldhaber and Hansen 2009) to ensure that a proctor-student match is actually also a teacher-student match.
5. There is no universally accepted method for estimating teacher effectiveness (Kane and Staiger 2008; Rothstein 2009). However, as we show in Goldhaber and Hansen (2009), the findings we report do not appear sensitive to the empirical specification of the VAM. For example, equation 1 is estimated without school- or classroom-level variables, but alternative specifications show that these variables explain only a very small proportion of teacher effectiveness and, therefore, teacher performance estimates are little influenced by including other school- or classroom-level variables in the model. For example, the correlation between the VAM teacher effects estimated with and without a school fixed effect in the model is over 0.9.
6. A teacher's tenure status is not observed in the data, but imputed, given the presence of a teacher in the same school district for four consecutive years.
7. We report the results from models that use the empirical Bayes "shrunken" teacher effectiveness estimates (McCaffrey et al. 2009), but the findings differ little if the unadjusted effects are used instead.
8. If we restrict the sample to just teachers in their fifth year, the pattern of results is similar to those reported in table 1. Similarly, the results differ very little when we use the first three years of teacher classroom performance to estimate effects rather than two.
9. Both student achievement and the teacher effect estimates included in the regressions are standardized by grade and year to zero mean and unit variance so the point estimates show the estimated effect size of a 1 standard deviation of prior teacher effectiveness (measured in student achievement terms) on current student achievement.
10. Note that, due to attrition, the number of unique teachers in the sample drops from 609 to 525 for this exercise.
11. There are 10,127 total predictions or about 100 per percentile.
12. A truncation of the bottom quarter of the teacher workforce is not as large a reduction as might appear at first blush since early-career teacher attrition is relatively high and many teachers who leave of their own accord are in the lowest quartile of performance (Goldhaber and Hansen 2009).

ACKNOWLEDGMENTS

The authors are grateful to Philip Sylling and Stephanie Liddle for research assistance. This brief also benefitted from helpful comments and feedback by Dale Ballou, Cory Koedel, Hamilton Lankford, Austin Nichols, Jesse Rothstein, Daniel McCaffrey, and Tim Sass, and helpful comments from participants at the Association for Public Policy Analysis and Management 2009 Fall Research Conference, the University of Virginia's 2008 Curry Education Research Lectureship Series, and the 2008 Economics Department Seminar Series at Western Washington University. The research presented here is based primarily on confidential data from the North Carolina Education Research Data Center at Duke University, directed by Clara Muschkin and supported by the Spencer Foundation. The authors wish to acknowledge the North Carolina Department of Public Instruction for its role in collecting this information and making it available. They gratefully acknowledge the Institute of Educational Studies at the Department of Education for providing financial support for this project. The views expressed in this brief do not necessarily reflect those of the University of Washington or the study's sponsor.

REFERENCES

- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007. "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Economics of Education Review* 26(6): 673–82.
- Goldhaber, Dan D. 2007. "Everyone's Doing It, but What Does Teacher Testing Tell Us about Teacher Effectiveness?" *Journal of Human Resources* 42(4): 765–94.
- Goldhaber, Dan D., and Michael Hansen. 2009. "Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions." Working Paper 2009-2. Seattle, WA: Center on Reinventing Public Education.
- Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger. 2006. "Identifying Effective Teachers Using Performance on the Job." Hamilton Project White Paper 2006-01. Washington, DC: The Brookings Institution.
- Hanushek, Eric A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19(2):141–64.
- . 2009. "Teacher Deselection." In *Creating a New Teaching Profession*, edited by Dan Goldhaber and Jane Hannaway (165–80). Washington, DC: Urban Institute Press.
- Joyce, Robert P. 2000. *The Law of Employment in North Carolina's Public Schools*. Chapel Hill: Institute of Government, University of North Carolina at Chapel Hill.
- Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." Working Paper 14607. Cambridge, MA: National Bureau of Economic Research.
- Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27(6): 615–31.
- McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood, and Kata Mihaly. 2009. "The Intertemporal Variability of Teacher Effect Estimates." *Education Finance and Policy* 4(4): 572–606.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73(2): 417–58.
- Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Students' Achievement: Evidence from Panel Data." *American Economic Review* 94(2): 247–52.
- Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125(1): 175–214.
- Weisburg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling. 2009. "The Widget Effect." *Education Digest* 75(2): 31–35.

ABOUT THE AUTHORS

Dan Goldhaber is professor in interdisciplinary arts and sciences at the University of Washington, Bothell; an affiliated scholar at the Urban Institute; and a coeditor of *Education Finance and Policy*. He leads the CALDER Washington team and is a member of CALDER's strategic planning group. He previously served as an elected member of the Alexandria City (Va.) School Board from 1997 to 2002. His work focuses on K-12 educational productivity and reform and addresses the role that teacher pay structure plays in teacher recruitment and retention as well as the influence of human resource practices on teacher turnover and quality. Dr. Goldhaber's work has been published in leading journals and has appeared in major media outlets such as National Public Radio, *Education Week*, *Washington Post*, and *USA Today*. His full C.V. is available at www.caldercenter.org.

Michael Hansen is a research associate with the Education Policy Center at the Urban Institute and an affiliated researcher with CALDER. His research interests include the economics of education, applied microeconomics, and labor economics. He previously worked as a research assistant with the Center on Reinventing Public Education in Seattle. Dr. Hansen was a 2009 finalist in the National Council on Teacher Quality Research Competition. He has researched teacher effort, value-added measurement, district-level contract choice for teachers, and national board certification and license testing for teachers. His full C.V. is available at www.caldercenter.org.

THE URBAN INSTITUTE
2100 M Street, N.W.
Washington, D.C. 20037



Phone: 202-833-7200
Fax: 202-467-5775
<http://www.urban.org>

Return service requested

Nonprofit Org.
US Postage PAID
Easton, MD
Permit No. 8098

National Center for Analysis of Longitudinal Data in Education Research

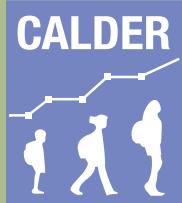
IN THIS ISSUE

Using Performance on the Job to Inform Teacher Tenure Decisions

To download this document, visit our web site,
<http://www.urban.org>.

For media inquiries, please contact
paffairs@urban.org.

National Center for Analysis of Longitudinal Data in Education Research



This research is part of the activities of the National Center for the Analysis of Longitudinal Data in Education Research (CALDER). CALDER is supported by Grant R305A060018 to the Urban Institute from the Institute of Education Sciences, U.S. Department of Education. More information on CALDER is available at <http://www.caldercenter.org>.

The paper on which this brief is based was presented at the annual meeting of the American Economic Association in Atlanta, Georgia, in January 2010. It, along with other papers presented in the session "Implicit Measurement of Teacher Quality," is published in The American Economic Review 100(2), May 2010.

The views expressed are those of the authors and do not necessarily reflect the views of the Urban Institute, its board, its funders, or other authors in this series. Permission is granted for reproduction of this document, with attribution to the Urban Institute.

