

Working PAPER

BY MARIESA HERRMANN, ELIAS WALSH, ERIC ISENBERG, AND ALEXANDRA RESCH

Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels

April 2013

Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels

Mariesa Herrmann, Elias Walsh, Eric Isenberg, and Alexandra Resch

Mathematica Policy Research

April 12, 2013

ABSTRACT

It is common in the implementation of teacher accountability systems to use a procedure known as Empirical Bayes shrinkage to adjust the teacher value-added estimates by their level of precision. Because value-added estimates based on fewer students and students with “hard-to-predict” achievement will be less precise, the procedure could have differential impacts on the probability that the teachers of fewer students or students with hard-to-predict achievement will be assigned consequences. This paper investigates how shrinkage affects the value-added estimates of teachers of hard-to-predict students, focusing on the context of threshold-based accountability systems. We found that the achievement of particular groups of students—students with low prior achievement and who receive free lunch—is harder to predict. Teachers of these types of students tend to have less precise value-added estimates, and shrinkage increases their estimates’ precision. Shrinkage also reduces the absolute value of the value-added estimates of teachers of hard-to-predict students. However, in our data, we found that shrinkage has no statistically significant effect on the relative probability that teachers of hard-to-predict students receive consequences.

I. INTRODUCTION

Due to the incentives provided by the federal Race to the Top program, districts and states are rapidly implementing new teacher evaluation systems to make high-stakes decisions about tenure, pay, and retention, based in part on statistical measures of teacher effectiveness. These evaluation systems generally incorporate multiple measures of effective teaching, such as student achievement growth, classroom observations, and student learning objectives (Mihaly et al. 2013). One of the most controversial has been the student achievement growth component, for which districts have used different approaches. One common approach is to use a value-added model (VAM) to estimate teachers' contributions to student achievement on standardized assessments.

VAMs predict individual student achievement based on the student's characteristics, including baseline achievement, and compare this prediction with the actual achievement of a teacher's students. The prediction is derived using data on other students in the state or district and represents what we would expect the student to achieve if he or she were taught by the average teacher. The difference between how a teacher's students actually performed and how they were predicted to perform represents the estimate of the teacher's value added to student achievement.

The precision of the teacher's value-added estimate is related to the average differences between the actual and predicted performance of the students in a teacher's class. Precision depends both on the number of students matched to the teacher and the accuracy of the model for predicting the achievement of the specific students in the teacher's class. All else being equal, the estimates of teachers matched to fewer students are more likely to be imprecise because one or two students' having unusually high or low achievement growth can more heavily influence these estimates. The accuracy of the model is affected by the existence of factors that influence achievement but are not observed in the data. Some students may have achievement that is hard to predict in the sense that their background characteristics are not very informative about their future achievement. The role of this potential *heteroskedasticity*—variation in the accuracy of predictions across student groups—has received relatively little attention in the value-added literature.¹ However, Stacy et al. (2012) found that students with low socioeconomic status and low prior achievement had less-accurate predicted scores than more-advantaged students. They showed that the value-added estimates of teachers who had many disadvantaged students were less stable over time.

It is common in the implementation of teacher accountability systems—including in the District of Columbia, New York City, and Los Angeles—to use a procedure known as Empirical Bayes (EB) shrinkage to adjust the teacher value-added estimates by their level of precision.² The adjusted

¹ A number of studies have examined the extent to which nonrandom assignment of students to teachers causes bias in teacher value-added estimates and teachers to be misclassified (for example, Aaronson et al. 2007; Kane and Staiger 2008; Rothstein 2010; Chetty et al. 2011; Guarino et al. 2012). Although these studies tend to focus on the point estimates, this paper examines the estimates' variances.

² The shrinkage estimates are called the best linear unbiased predictors (BLUPs) because they minimize the expected mean squared error (MSE) between teachers' estimates and their actual contributions. However, critics of the approach note that it obtains these properties by intentionally introducing statistical bias in the estimates and does not generally provide an accurate ranking of teachers (Tate 2004).

estimate is a weighted average of the teacher's initial value-added estimate and the value-added of an average teacher, with more precise estimates receiving greater weight. This procedure is called shrinkage because less precise estimates receive lower weight in the adjustment and are shrunk toward the average. Because value-added estimates based on fewer students and students with hard-to-predict achievement will be less precise, the procedure could have differential impacts on the probability that the teachers of fewer students or students with hard-to-predict achievement will be assigned consequences. For example, in the absence of shrinkage, teachers with very few students could be more likely to end up in the extremes of the distribution.³ Similarly, if teachers of disadvantaged students have imprecise initial value-added estimates, shrinkage could reduce the probability that these teachers end up in the extremes of the distribution.

How shrinkage affects the probability that teachers of particular types of students will receive consequences depends on the designs of accountability systems. As shown in Table 1, most evaluation systems use thresholds to identify highly effective or ineffective teachers and assign consequences. Because consequences are often discrete—for example, whether to retain the teacher or whether to require the teacher to receive professional development—very small changes in a teacher's estimate near the threshold can have substantial consequences. In designing a high-stakes accountability system, one goal is to avoid classification errors.⁴

In these systems, shrinkage is conceptually appealing because it produces estimates of teacher effectiveness that are conservative in the assignment of consequences to teachers. Shrinkage estimates are conservative because an imprecise estimate will be shrunk more to the overall average, and the teacher might be less likely to be assigned consequences as a result. Because precision depends on the number of students and the characteristics of the students in a teacher's class, shrinkage could potentially reduce the probability that teachers of these types of students receive consequences. This might be desirable in evaluation systems because differences in teachers' probabilities of being misclassified could be considered unfair and have deleterious effects on teachers' incentives to teach classes that include certain groups of students.

This paper investigates how shrinkage affects the value-added estimates of teachers of hard-to-predict students, focusing on the context of threshold-based accountability systems. We examine three main questions:

1. What student characteristics are correlated with having hard-to-predict test scores?
2. Are these student characteristics correlated with how much teacher value-added estimates change when shrinkage methods are used?
3. Are these student characteristics correlated with how much shrinkage affects the probability that these teachers' estimates exceed policy-relevant thresholds?

Consistent with heteroskedasticity, we find that the achievement of particular students, such as students with low prior achievement and who receive free lunch, is harder to predict. Teachers of

³ Kane and Staiger (2002) found that small schools were more likely to have large changes in achievement across years than large schools.

⁴ Shrinkage could also differentially reduce the probability of consequences for teachers with greater variation in their true impacts on students. Variation in true impacts cannot be disentangled from variation due to random shocks to student performance, and the shrinkage procedure errs on the side of attributing imprecision to random variation rather than to variation in true impacts.

these types of students tend to have less precise value-added estimates and shrinkage increases their estimates' precision. Shrinkage also reduces the absolute value of the value-added estimates of teachers of hard-to-predict students. However, in our data, shrinkage has no statistically significant effect on the relative probability that teachers of hard-to-predict students receive consequences.

The main contribution of this paper is to examine heteroskedasticity and shrinkage in the context of accountability systems based on thresholds. This differs from other papers on shrinkage, which examine how shrinkage affects teacher ranks (Tate 2004; Guarino et al. 2012), and other work on heteroskedasticity, which examines its effect on the precision and inter-year stability of value-added estimates (Stacy et al. 2012). Our purpose is to examine how shrinkage affects the probability that teachers of hard-to-predict students are classified at the extremes of the value-added distribution of teacher effectiveness, because many evaluation systems use these thresholds to determine consequences.

II. THEORY

To illustrate how heteroskedasticity affects estimates of teacher value added, we present the following VAM:

$$(1) Y_{igt} = \lambda Y_{i,g-1} + \alpha' \mathbf{X}_{igt} + \theta_t + \varepsilon_{igt}$$

where Y_{igt} is the test score of student i in grade g who has teacher t and $Y_{i,g-1}$ is the prior test score for student i in the previous grade $g-1$. The vector \mathbf{X}_{igt} denotes control variables for student demographic characteristics, θ_t is a vector of teacher effects,⁵ and ε_{igt} is the student-level error term.

Under homoskedasticity, all student-level errors ε_{igt} have the same variance—the accuracy of a predicted score does not depend on the characteristics of the student. However, estimates of precision typically allow for the possibility of heteroskedastic errors. One possible cause of heteroskedasticity is measurement error in test scores. Typically, standardized test scores near the center of the distribution are more reliable than scores in the tails of the distribution (for example, CTB/McGraw-Hill 2011).⁶ Consequently, students who are likely to be at the top or the bottom of achievement distribution might tend to have less-accurate predicted scores. Another reason for heteroskedasticity is the presence of unobserved factors that affect achievement. For example, we do not observe parental involvement or students' motivation, which could affect achievement and vary across different types of students.

⁵ Teacher effects can be estimated as random variables that are independent of other factors affecting test scores or as fixed effects that might be correlated with the other factors affecting test scores. We use the latter method, which estimates teacher effects as fixed effects, to account for the possibility that teacher assignments are correlated with the covariates included in the model.

⁶ Standardized tests are designed to differentiate students who are proficient from those who are not, so they provide more reliable scores for students around the level of proficiency. Koedel et al. (2012) developed a method for accounting for measurement error that varies across the achievement distribution.

Shrinkage places less weight on imprecise initial value-added estimates when constructing the final value-added estimates. If teachers who are assigned students with large residuals tend to have imprecise estimates of initial value added, then the shrinkage procedure will proportionately shrink the estimates of these teachers more and the final estimates of these teachers will be proportionately more precise than without shrinkage. This relationship means that shrinkage reduces but does not eliminate differences in variances of value-added estimates.

To see this, consider the EB shrinkage procedure outlined in Morris (1983). A teacher's shrinkage estimate is the weighted average of the teacher's pre-shrinkage value-added estimate and the estimate for the average teacher. Let $\hat{\theta}_t$ be the mean-centered pre-shrinkage point estimate for teacher t from the value-added regression model, and let the associated heteroskedasticity-robust variance estimate be $\widehat{Var}[\hat{\theta}_t] = \hat{\sigma}_t^2$. The EB estimate for a teacher is approximately equal to a precision-weighted average of $\hat{\theta}_t$ and the overall mean of all estimated teacher effects.⁷ Because the overall mean of the estimated teacher effects is zero by design, the teacher's EB estimate $\hat{\theta}_t^{EB}$ can be written as follows:

$$(2) \quad \hat{\theta}_t^{EB} \approx \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\sigma}_t^2} \right) \hat{\theta}_t$$

where $\hat{\sigma}$ is an estimate of the standard deviation of teacher effects (purged of sampling error), which is constant for all teachers. The term $\hat{\sigma}^2 / (\hat{\sigma}^2 + \hat{\sigma}_t^2)$ must be less than one, so the EB estimate always has a smaller absolute value than the initial estimate.

Although heteroskedasticity in student background characteristics does not affect the pre-shrinkage point estimates, it does affect the variance of the teacher estimates and, thus, the EB estimates. The change in the EB estimate in response to a change in the variance of the pre-shrinkage estimate is given by the following:

$$(3) \quad \frac{\partial \hat{\theta}_t^{EB}}{\partial \hat{\sigma}_t^2} = -\frac{\hat{w}_t^2}{\hat{\sigma}^2} \hat{\theta}_t$$

where $\hat{w}_t \equiv \hat{\sigma}^2 / (\hat{\sigma}^2 + \hat{\sigma}_t^2)$, the weight on the pre-shrinkage estimate. The negative sign indicates that estimates of teachers who teach students with characteristics that produce large residuals are shrunk more than those of teachers who teach students with smaller residuals. The extra shrinkage is largest for teachers with larger unshrunk estimates and for teachers with more of the types of students who tend to produce large residuals. The extra shrinkage will tend to distort the relationship between teacher effectiveness and teachers' average student characteristics related to

⁷ In Morris (1983), the EB estimate does not exactly equal the precision-weighted average of the two values due to a correction for bias. This adjustment increases the weight on the overall mean by $(K - 3)/(K - 1)$, where K is the number of teachers. For ease of exposition, we have omitted this correction from the description given here and consider this expression to hold with equality.

heteroskedasticity (or other characteristics of teachers or classrooms related to heteroskedasticity). This is an example of the trade-off implicit in shrinkage between introducing bias and the choice to minimize the mean squared error (MSE) of the estimates;⁸ the negative relationship in Equation (3) does not indicate that these teachers are more or less effective on average—it is instead an artifact of being conservative in handing out extreme value-added scores to teachers.

One benefit of the trade-off between bias and MSE is the increased efficiency of the EB estimator. This can be seen in the lower variances associated with the EB estimates. Ignoring the sampling error associated with $\hat{\sigma}^2$ and $\hat{\sigma}_t^2$, the variance of a teacher's EB estimate is approximately equal to⁹

$$(4) \quad \overline{\text{Var}}[\hat{\theta}_t^{EB}] \approx \hat{w}_t \hat{\sigma}_t^2$$

Thus, the EB estimate has a smaller variance than that of the pre-shrinkage estimate, because \hat{w}_t is less than one.

Heteroskedasticity affects $\overline{\text{Var}}[\hat{\theta}_t^{EB}]$ through the variance on the pre-shrinkage estimates. The change in the variance of a teacher's EB estimate in response to a change in the variance of the pre-shrinkage estimate is proportional to the square of the weight on the point estimate in the EB calculation:

$$(5) \quad \frac{\partial \overline{\text{Var}}[\hat{\theta}_t^{EB}]}{\partial \hat{\sigma}_t^2} = \hat{w}_t^2$$

Because \hat{w}_t is less than one but greater than zero, shrinkage reduces but does not eliminate heteroskedasticity from the EB estimates. This suggests that shrinkage might mitigate but will not eliminate the differential probability that teachers of hard-to-predict students are classified in the extremes of the value-added distribution.

III. VALUE-ADDED MODEL AND DATA

We estimate teacher value added using data from the District of Columbia Public Schools (DCPS) and the Office of the State Superintendent of Education of the District of Columbia

⁸ Morris (1983) and others refer to EB estimates as unbiased. For example, the EB random effects estimator is often called the BLUP. The EB estimates are unbiased only in the sense that the expectation of the mean over all groups (teachers) gives an unbiased estimate of the true mean over the population of group effects. As a direct consequence of Equation (3), a particular teacher's EB estimate is not generally an unbiased estimate of that teacher's true effect.

⁹ The actual formula we employ for the variance of the EB estimates is based on a formula in Morris (1983) that accounts for sampling error in the estimated variances and makes degrees of freedom corrections. For ease of exposition, we have omitted these corrections from the description given here and consider this expression to hold with equality.

(OSSE).¹⁰ The data contain information on students' test scores and demographic characteristics and, importantly for the estimation of value added, enable students to be linked to their teachers. In the DCPS, students in grades 3 through 8 and 10 are tested in math and reading using the District of Columbia Comprehensive Assessment System (DC CAS) tests. Our analysis is based on students who were in grades 4 through 8 during the 2011–2012 school year and who had both pre- and post-test scores. To enable us to compare students across grades, we standardize student test scores within subject, year, and grade to have a mean of 0 and a standard deviation of 1. We exclude students who are repeating the grade so that in each grade, we compare only students who completed the same examinations. We estimate the following VAM separately for teachers of elementary (grades 4 and 5) and middle school (grades 6 through 8) students:

$$(6) \quad Y_{ig} = \lambda_g Y_{i(g-1)} + \omega_g Z_{i(g-1)} + \alpha' X_i + \delta_g + \eta' T_i + \varepsilon_{ig}$$

where Y_{ig} is the post-test score for student i in grade g and $Y_{i(g-1)}$ is the same-subject pre-test for student i in grade $g-1$ during the previous year. The variable $Z_{i(g-1)}$ denotes the pre-test in the opposite subject. Thus, when estimating teacher effectiveness in math, Y_{ig} and $Y_{i(g-1)}$ represent math tests, with $Z_{i(g-1)}$ representing reading tests and vice versa. The pre-test scores capture prior inputs into student achievement, and the associated coefficients λ_g and ω_g vary by grade. The vector X_i denotes control variables for individual student background characteristics—specifically, indicators for free lunch eligibility, reduced-price lunch eligibility, limited English proficiency, and special education status. The coefficients on these characteristics are constrained to be the same across grades within each grade span. The vector δ_g includes grade indicators.

The vector T_i contains one indicator variable for each teacher. A student contributes one observation to the model for each teacher to whom the student is linked. This contribution is based on a roster confirmation process that enables teachers to indicate whether and for how long they have taught the students on their administrative rosters and to add any students who are not listed on their administrative rosters (Hock and Isenberg 2012). Students are weighted in the regression according to their dosage, which indicates the amount of time the teacher taught the student.¹¹ The vector η includes one coefficient for each teacher. Finally, ε_{ig} is the random error term.

The VAM in Equation (6) is estimated in two regression steps:

¹⁰ Isenberg and Hock (2012) provide additional information about the DC data and VAM used for accountability purposes. The VAM used in this paper differs slightly.

¹¹ To estimate the effectiveness of teachers who share students, we use a technique called the full roster plus method, which attributes equal credit to teachers of shared students. In this method, each student contributes one observation to the model for each teacher to whom he or she is linked and students are weighted based on the dosage they contribute (Hock and Isenberg 2012). Then we create additional observations to equalize the dosage that each student contributes to the model. The proportion of dosage each student contributes to a teacher remains unchanged because the additional observations are linked to a distinct set of so-called shadow teacher indicators. Coefficient estimates for these shadow teachers are discarded.

1. **Measurement error correction.** Measurement error in the pre-test scores will attenuate the estimated relationship between the pre- and post-test scores. We adjust for measurement error using an errors-in-variables correction (`eivreg` in Stata) that relies on published information on the test-retest reliability of the DC CAS. We use errors-in-variables regression to regress the post-test score on the pre-test scores, student background characteristics, and grade and teacher indicators. Because the errors-in-variables regression does not allow for standard errors to be clustered by student, in this step we obtain adjusted gain scores, which subtract the predicted effects of the pre-test scores from the post-test scores, which we use to obtain the initial teacher effects in the next step. Notably, because the measurement error correction accounts only for measurement error that is constant across the test score distribution, any measurement error that varies across the distribution will contribute to heteroskedasticity.
2. **Main regression.** We estimate teacher effects by regressing the adjusted gain scores from the first step on student background characteristics, grade and teacher indicators, and by clustering standard errors by student. The initial teacher value-added estimates are the coefficients on the teacher indicators in this regression, and their variance is given by the squared standard errors of the coefficient estimates. As a final step, we remove any teachers with fewer than 15 students from the teacher model and recenter the value-added estimates to have a mean of 0.

We then calculate the EB estimates by applying Equation (2) to the initial estimates and variances from the main value-added regression. We also calculate the variance of the EB estimates using Equation (4).

Teachers vary in the characteristics of students that they teach. Table 2 displays the summary statistics for the 297 elementary school teachers (top panel) and 182 middle school teachers (bottom panel) in our sample. For simplicity, we restrict our analysis to math teachers and exclude teachers linked to fewer than 15 or more than 150 students.

Because groups of students with large residuals will affect the precision of teacher value-added estimates only if they are unevenly distributed across teachers, the table reports the teacher-level averages of student characteristics. For homeroom teachers—typically elementary school teachers—these values are the average characteristics of the students in the teachers’ classrooms, whereas for departmentalized teachers, these values are averaged across multiple classes. Table 2 shows that there is wide variation in students’ test scores and classroom composition across teachers. For example, in math, the average pre-test scores of teachers’ students range from -1.34 to 1.36 standard deviations in elementary school and from -1.12 to 1.16 standard deviations in middle school. Teachers also vary substantially in the percentage of students eligible for free and reduced-price lunch, limited English proficiency, and special education students they teach.

Table 2 also reports the number of student-equivalents taught by teachers, which weights the number of students taught by teachers by their dosages—that is, the portions of time that individual students were enrolled in their teachers’ classes. Elementary school teachers are typically linked to fewer students than middle school teachers, consistent with more departmentalized teachers in middle school. However, even within a grade span there is considerable variation in student-equivalents across teachers, which could affect the precision of value-added estimates.

Post-shrinkage value-added estimates and variances are smaller than the pre-shrinkage estimates. Table 2 presents the pre-shrinkage and shrinkage value-added estimates and their estimated variances. For elementary school teachers, the pre-shrinkage estimates range from -0.90 to

0.96 and the estimated pre-shrinkage variances range from 0 to 0.07. Because the shrinkage procedure assigns the pre-shrinkage estimates and variances weights that are less than 1—the average weight for elementary teachers is 0.85—the shrinkage estimates have a smaller range than the pre-shrinkage estimates and the shrinkage variances are smaller than the pre-shrinkage variances. For elementary school teachers, the post-shrinkage estimates range from -0.77 to 0.71 and the estimated post-shrinkage variances range from 0.04 to less than 0.005.

In the next section, we describe our methodology to examine how differences in the characteristics of teachers' students are related to precision and how accounting for imprecision using shrinkage affects the probability that teachers are classified in the extremes of the value-added distribution based on their students' characteristics.

IV. EMPIRICAL APPROACH

To describe heteroskedasticity in the student-level errors, we estimate the relationship between the squared residuals from the VAM—a measure of the accuracy of the prediction—and student characteristics using the following student-teacher-level regression:

$$(7) \quad \hat{\varepsilon}_{tig}^2 = \alpha + \beta X_i + v_{tig}$$

The outcome $\hat{\varepsilon}_{tig}^2$ is the squared residuals for student i in grade g with teacher t from the student-teacher-level regression of student test scores on student characteristics and grade and teacher indicators. Because the VAM contains teacher indicators, these residuals are the remaining variation in student outcomes after taking into account the average differences in the effectiveness of students' teachers. Thus, unpredictable outcomes for certain types of students do not simply reflect assignment of these students to teachers with a wider distribution of ability levels. Instead, the residuals measure the remaining difference between these students' outcomes and those of other students taught by the teacher. The vector X_i denotes student characteristics (that is, average math and reading pre-test scores, eligibility for free or reduced-price lunches, limited English proficiency, and special education status), and v_{igt} is an error term.

The regression in Equation (7) is intended to provide a parsimonious description of heteroskedasticity and not a formal test for its presence. For ease of interpretation, we estimate these regressions separately for each student characteristic and grade span; the sole exceptions to the separate regressions are eligibility for free or reduced-price lunch, which are entered simultaneously into the regression.¹² To address repeated observations of students, we estimate standard errors that account for clustering in the residual at the student level and weight the regressions by the student-teacher dosages.

After examining which types of students tend to have higher residuals, we analyze the relationship between these student characteristics and the precision and magnitude of their teachers' pre- and post-shrinkage value-added estimates. We estimate the following teacher-level bivariate regression model, again using separate regressions for each student characteristic and grade span:

$$(8) \quad W_t = \alpha + \beta X_t + \varepsilon_t$$

¹² We do not examine higher-order polynomials for pre-test scores.

where W_t represents a measure of the precision or magnitude of the pre- or post-shrinkage value-added estimate for teacher t . Specifically, for both pre- and post-shrinkage estimates, we examine as outcomes the natural log of the estimates' variance, the absolute value of the estimates, and indicators for the estimates being above or below a given percentile of the value-added distribution. The vector X_t represents the average characteristics of the teacher's students from Table 2, including measures of the student-equivalents taught by the teacher—the natural log of the total or the raw total. As we did in the student-teacher-level regressions, for ease of interpretation, we estimate separate regressions for each student characteristic (with the exception of eligibility for free or reduced-price lunch) and grade span. We calculate robust standard errors for each teacher.

V. RESULTS

Table 3 presents the results from the regressions that examine the relationship between student characteristics and the various outcomes. The first column displays the results from the student-teacher-level regressions of the squared residuals on student characteristics and provides evidence that some groups of students have significantly larger residuals than others. This is consistent with heteroskedasticity and shows which types of students have achievement that is hard to predict. In elementary school, the hard-to-predict students are those who are eligible for free lunch, receive special education services, and have lower pre-test scores.¹³ The same characteristics are also significantly associated with larger residuals in middle school, as is limited English proficiency. These results are consistent with those of Stacy et al. (2012), who found higher residuals for students eligible for free lunch and students with lower prior test scores, but do not explicitly examine special education or limited English proficiency status.

Consistent with the significant relationships between particular student characteristics and the student-level residuals, elementary school teachers assigned hard-to-predict students also have less precise value-added estimates. In column 2 of Table 3, we provide the results from teacher-level regressions, which separately regress the natural log of teachers' pre-shrinkage variances on each of the characteristics of teachers' students and the natural log of the teachers' number of student-equivalents.¹⁴ For elementary school teachers, a one-standard-deviation decrease in students' average pre-test scores in math is significantly associated with a 20 percent increase in the teacher's pre-shrinkage variance. Lower pre-test scores in reading and free lunch eligibility are also significantly associated with higher pre-shrinkage variances in elementary school. Interestingly, limited English proficiency is significantly correlated with *lower* pre-shrinkage variances in elementary school, and

¹³ To determine whether students with very high or very low pre-test scores have larger residuals, as would be expected from measurement error in the tests, we also examined the relationship between the residuals and the absolute value of the pre-test scores, an indicator for having a below-mean pre-test score, and an interaction between the absolute value of the pre-test score and having a below-mean pre-test score (results not shown). Although having a pre-test score at either extreme of the distribution is associated with larger residuals, residuals are significantly higher for students at the bottom of the pre-test distribution than for those at the top.

¹⁴ We use the natural log of the variance as the dependent variable in these specifications for two reasons. First, the natural log of the variance more closely meets the normality assumption for efficient ordinary least squares (OLS) estimates. Second, the specification reflects the theoretical relationship between the variance and observable characteristics: $\ln(\hat{V}_i) = \ln(\hat{\sigma}_i^2) - \ln(N_i) = \ln(\alpha + \beta X_i + \varepsilon_i) - \ln(N_i)$, where \hat{V}_i is the estimated variance and N_i is the number of student equivalents.

there is no significant relationship between the pre-shrinkage variances and special education status. Having a larger number of students is associated with lower pre-shrinkage variances.¹⁵

The same general patterns hold in middle school. Lower pre-test scores in math and reading, free lunch eligibility, and special education status are significantly associated with higher pre-shrinkage variances. The magnitudes of the associations are generally larger for middle school teachers than for elementary school teachers. For example, a one standard deviation decrease in students' math scores is significantly associated with a 64 percent increase in a teacher's pre-shrinkage variance in the middle school grades. Column 2 of Table 3 also investigates the relationship between the natural log of the teachers' number of student-equivalents and the variance.

These same relationships with student characteristics are significantly smaller for post-shrinkage variances. Recall that shrinkage increases the precision of estimates at the cost of intentionally introducing bias. This effect of shrinkage is reflected in columns 3 and 4 of Table 3. In column 3, we examine the relationships among teachers' post-shrinkage variances, the characteristics of their students, and teachers' number of student-equivalents. Compared with the pre-shrinkage variances, the post-shrinkage variances have the same number of statistically significant relationships with student characteristics, although the size of these correlations is significantly smaller for the post-shrinkage variances. For example, for elementary school students, a one standard deviation decrease in students' average pre-test scores in math is significantly associated with a 17 percent increase in the teacher's post-shrinkage variance, compared with a 20 percent increase in the pre-shrinkage variance, a statistically significant difference. Statistical tests of the differences between the coefficients for pre-shrinkage and post-shrinkage models (columns 2 and 3, respectively) can be seen in column 4. Similarly, shrinkage also reduces the size of the relationships between free lunch eligibility and teachers' number of student-equivalents and the estimates' variance.

Shrinkage also affects the absolute value of the estimates of teachers with hard-to-predict students. Column 5 of Table 3 presents results from teacher-level regressions of the absolute value of teachers' pre-shrinkage value-added estimates on student characteristics and teachers' number of student-equivalents. In elementary school, the pre-shrinkage value-added estimates for teachers of students with lower pre-test scores and teachers with more students eligible for free lunch are significantly further from the overall teacher mean relative to the pre-shrinkage value-added estimates of other teachers.¹⁶ In middle school, student characteristics do not have a significant relationship with the absolute value of the pre-shrinkage value-added estimates, possibly because we lack the power to precisely estimate these relationships. In column 6, we examine the relationship between the absolute value of teachers' post-shrinkage value-added estimates, student characteristics, and teachers' number of student-equivalents. As before, student characteristics and teachers' number of student-equivalents still have statistically significant relationships with the absolute values of the

¹⁵ The relationship coefficient on the natural log of the number of student equivalents is not -1 , as would be expected from the equation in the previous footnote. This is likely due to correlations between class size and other student characteristics that are correlated with teachers' variances.

¹⁶ The absolute value is measured relative to the overall teacher mean, rather than the mean conditional on the characteristics of teachers' students. The pre-shrinkage estimates could be related to the characteristics of teachers' students if teachers are not equitably distributed. The results in columns 5 through 7 of Table 3 do not adjust for correlated teacher assignments. However, shrinkage is the only difference between the value-added estimates used in columns 5 and 6, so the difference in the relationships between student characteristics and the absolute values of the value-added estimates (column 7) can be attributed to shrinkage.

shrinkage estimates, but these relationships are weaker than their relationships with the pre-shrinkage absolute values were.

We next assess how shrinkage affects the probability that teachers are assigned positive or negative consequences. We focus on the extremes of the distribution because teachers with such estimates are the most likely to be targeted for rewards or sanctions.¹⁷ In particular, we examine the relationship between student characteristics and indicator variables for whether the estimates fell into the top or bottom deciles of their distributions.¹⁸ We presume that districts “grade on a curve”—that is, they set thresholds for categories of effectiveness based on teachers’ positions in the value-added distribution. By using the same relative thresholds of effectiveness for pre- and post-shrinkage estimates, we use different absolute thresholds as measured in standard deviations of teacher value added. Examining value-added estimates in a relative rather than absolute sense also affects how to think about how estimates change after shrinkage. In an absolute sense, all estimates move toward the center of the distribution, with less precise estimates moving more. In relative terms, however, relatively imprecise initial estimates will move toward the center, and relatively precise estimates will move toward the extremes only if they change places in a rank order with imprecise initial estimates.

We present these results in Table 4. In column 1, using a linear probability model, we regress an indicator variable for the pre-shrinkage estimate falling in either the top or bottom deciles of the pre-shrinkage value-added distribution on each student characteristic and on teachers’ number of student-equivalents. For elementary school, we find that having students with low pre-test scores, students eligible for free lunch, or smaller numbers of students increases the probability that teachers’ pre-shrinkage estimates are in either the top or bottom deciles of the distribution. In middle school, there is no relationship between the characteristics of teachers’ students and the probability that the estimates fall in the top or bottom deciles. Column 2 examines the probability that the post-shrinkage estimate is either in the top or bottom deciles of the post-shrinkage value-added distribution and column 3 tests whether the difference between the pre- and post-shrinkage relationships is statistically significant. We find that shrinkage has little effect on the probability that teachers of particular student groups are classified in the top or bottom deciles of the value-added distribution. Columns 4 through 6 repeat the same analysis using an indicator variable for the estimate falling in either the top or bottom quintiles of the distribution. Because the top and bottom quintiles cover a much larger range of teacher effectiveness, there are fewer significant relationships between student characteristics and the probability of the estimate’s being classified in either the top or bottom quintile, though the magnitude of some of the relationships appears to have increased. However, shrinkage does not significantly affect the probability that teachers of hard-to-predict students are classified in either the top or bottom quintiles of the value-added distribution. Shrinkage might make little difference in classifying teachers of hard-to-predict students when accountability thresholds are rescaled along with the estimates.

¹⁷ The probability that teachers in the top or bottom decile receive consequences depends on the weight that value-added estimates receive in the overall evaluation, the correlation between value-added estimates and other evaluation components, and the placement of the thresholds.

¹⁸ As a sensitivity analysis, we replace the top and bottom deciles by the top and bottom quintiles.

VI. CONCLUSION

Recent work has raised concerns about the effect of heteroskedasticity on estimates of value added because heteroskedasticity can reduce the precision of value-added estimates for teachers of particular types of students (for example, Stacy et al. 2012). Because shrinkage adjusts value-added estimates by their level of precision, it might be viewed as a method to mitigate heteroskedasticity. This paper empirically investigates how shrinkage affects the relationship between student characteristics and the precision of value-added estimates, the absolute value of those estimates, and the probability that the estimates fall in the extremes of the value-added distribution. Consistent with heteroskedasticity, we find that particular student characteristics—lower pre-test scores and eligibility for free lunch—are correlated with the precision of a teacher’s estimate. We show that shrinking the estimates improves the precision of the adjusted estimates and reduces the absolute value of estimates for teachers with hard-to-predict students. However, in these data, shrinking has no statistically significant impact on the relative probability of exceeding a threshold for teachers of hard-to-predict students.

REFERENCES

- Aaronson, Daniel, Lisa Barrow, and William Sander. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics*, vol. 25, no. 1, 2007, pp. 95–135.
- Chetty, Raj, John Friedman, and Jonah Rockoff. "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood." Working paper no. 17699. Cambridge, MA: National Bureau of Economic Research, 2011.
- CTB/McGraw-Hill. "Technical Report for Spring 2011 Operational Test Administration of DC CAS." Monterey, CA: CTB/McGraw-Hill, 2011.
- Guarino, Cassandra M., Michelle Maxfield, Mark D. Reckase, Paul Thompson, and Jeffrey M. Wooldridge. "An Evaluation of Empirical Bayes' Estimation of Value-Added Teacher Performance Measures." Working paper no. 31. East Lansing, MI: Education Policy Center at Michigan State University, 2012.
- Hock, Heinrich, and Eric Isenberg. "Methods for Accounting for Co-Teaching in Value-Added Models." Washington, DC: Mathematica Policy Research, June 2012.
- Isenberg, Eric, and Heinrich Hock. "Measuring School and Teacher Value Added in DC, 2011-2012 School Year." Final report submitted to the Office of the State Superintendent of Education for the District of Columbia and the District of Columbia Public Schools. Washington, DC: Mathematica Policy Research, August 31, 2012.
- Kane, Thomas J., and Douglas O. Staiger. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, vol. 16, no. 4, 2002, pp. 91–114.
- Kane, Thomas J., and Douglas O. Staiger. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." Working paper no. 14607. Cambridge, MA: National Bureau of Economic Research, 2008.
- Koedel, Cory, Rebecca Leatherman, and Eric Parsons. "Test Measurement Error and Inference from Value-Added Models." Unpublished manuscript. Columbia, MO: University of Missouri, 2012.
- Mihaly, Kata, Daniel F. McCaffrey, Douglas O. Staiger, and J.R. Lockwood. "A Composite Estimator of Effective Teaching." Seattle, WA: Bill & Melinda Gates Foundation, 2013.
- Morris, Carl N. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, vol. 78, no. 381, 1983, pp. 47–55.
- Rothstein, Jesse. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics*, vol. 125, no. 1, 2010, pp. 175–214.
- Stacy, Brian, Cassandra Guarino, Mark Recklase, and Jeffrey Wooldridge. "Does the Precision and Stability of Value-Added Estimates of Teacher Performance Depend on the Types of Students They Serve?" Preliminary draft. East Lansing, MI: Education Policy Center at Michigan State University, October 2012.

Tate, Richard L. "A Cautionary Note on Shrinkage Estimates of School and Teacher Effects."
Florida Journal of Educational Research, vol. 42, 2004, pp. 1–21.

Table 1. Teacher Accountability Systems

Evaluation System	State/District	Components	Levels of Performance and Consequences
Florida Teaching Evaluation System	Florida	Student performance (value-added), instructional performance (FEAP), professional and job responsibilities	Four levels of performance based on thresholds: highly effective, effective, needs improvement (or for teachers in the first 3 years, developing), and unsatisfactory. Performance levels for the value-added component are determined by whether the 95% confidence interval falls above the average (highly effective), contains the average (effective), or falls below the average (unsatisfactory).
IMPACT	District of Columbia Public Schools (Washington, DC)	50% student achievement (35% value-added and 15% teacher-assessed student data); 40% instructional expertise (classroom observations); and 10% commitment to school community (collaboration, professionalism)	Four levels of performance based on thresholds: highly effective, effective, developing, and minimally effective. Determine advancement to Leadership Initiative for Teachers (LIFT) career stage, bonuses, base salary increases, and separation.
Ohio Teacher Evaluation System	Ohio	50% teacher performance (professional growth or improvement plan, observations, walk-throughs, and conferences) and 50% student growth (teacher value-added combined with vendor or LEA measures)	Four levels of performance based on a matrix that maps three levels of student growth and four levels of teacher performance into overall ratings: accomplished, proficient, developing, and ineffective. Local boards of education will use the evaluation results for retention and promotion decisions and for dismissals.
Teacher Effectiveness (Phase II Implementation)	Pennsylvania	50% observation; 15% building-level data (PSSA achievement, PVAAS growth, graduation rate, promotion rate, attendance, AP course participation, and SAT/PSAT); 15% teacher-specific data (PSSA achievement, PVAAS growth, IEP growth, district rubrics); and 20% elective data	Four levels of performance based on thresholds: distinguished, proficient, needs improvement, and failing. Needs improvement or failing results in participation in a performance improvement plan.
Tennessee Educator Acceleration Model (TEAM)	Tennessee	50% observations and conferences (at least four evaluations annually); 35% TVAAS; and 15% achievement measure (TCAP and EOC evaluations)	Five levels of performance based on thresholds: significantly above expectations, above expectations, at expectations, below expectations, and significantly below expectations. Overall ratings determine eligibility for tenure and whether tenured teachers are returned to probationary status.

Source: Authors' research from state and district department of education websites.

Note: Value-added scores are not available for teachers in all grades and subjects, so the weights on components might differ for teachers who do not receive these scores.

AP = advanced placement; EOC = end of course; FEAP = Florida Educator Accomplished Practices; IEP = individualized education plan; LEA = local education agency; PSAT = Preliminary Scholastic Aptitude Test; PSSA = Pennsylvania System of School Assessment; PVAAS = Pennsylvania Value-Added Assessment System; SAT = Scholastic Assessment Test; TCAP = Tennessee Comprehensive Assessment Program; TVAAS = Tennessee Value-Added Assessment System.

Table 2. Teacher-Level Summary Statistics

Variables	Mean	Standard Deviation	Minimum	Maximum
<i>Elementary School</i>				
Math post-test score	0.02	0.57	-1.69	1.26
Math pre-test score	0.02	0.55	-1.34	1.36
Reading pre-test score	0.04	0.49	-1.21	1.25
Free lunch eligible	0.69	0.27	0.00	1.00
Reduced-price lunch eligible	0.06	0.06	0.00	0.23
Limited English proficiency	0.08	0.14	0.00	0.68
Special education	0.10	0.09	0.00	0.46
Grade 4	0.50	0.49	0.00	1.00
Grade 5	0.50	0.49	0.00	1.00
Student-equivalents	23.73	12.03	5.03	83.40
Pre-shrinkage value-added estimates	0.01	0.30	-0.90	0.96
Estimated variance of pre-shrinkage value-added estimates	0.01	0.01	0.00	0.07
Shrinkage value-added estimates	0.00	0.25	-0.77	0.71
Estimated variance of shrinkage value-added estimates	0.01	0.01	0.00	0.04
Number of teachers	297			
<i>Middle School</i>				
Math post-test score	-0.04	0.49	-1.27	1.17
Math pre-test score	-0.05	0.50	-1.12	1.16
Reading pre-test score	-0.04	0.44	-1.28	1.06
Free lunch eligible	0.72	0.21	0.00	0.98
Reduced-price lunch eligible	0.07	0.05	0.00	0.33
Limited English proficiency	0.08	0.13	0.00	0.58
Special education	0.13	0.10	0.00	0.92
Grade 6	0.39	0.45	0.00	1.00
Grade 7	0.30	0.38	0.00	1.00
Grade 8	0.31	0.40	0.00	1.00
Student-equivalents	54.07	25.54	7.99	138.75
Pre-shrinkage value-added estimates	-0.01	0.20	-0.58	0.63
Estimated variance of pre-shrinkage value-added estimates	0.01	0.01	0.00	0.04
Shrinkage value-added estimates	0.00	0.17	-0.40	0.58
Estimated variance of shrinkage value-added estimates	0.01	0.00	0.00	0.02
Number of teachers	182			

Source: Administrative data from OSSE and DCPS.

Note: Rows with student characteristics are based on teacher-level averages. Value added is estimated at the teacher level.

DCPS = District of Columbia Public Schools; OSSE = Office of the State Superintendent of Education of the District of Columbia.

Table 3. Effects of Student Characteristics on Teacher Value-Added Estimates

Characteristic	Log Variance of Individual Teacher Value-Added Estimates				Absolute Value of Individual Teacher Value-Added Estimates		
	Squared Residuals	Pre-Shrinkage	Post-Shrinkage	Difference	Pre-Shrinkage	Post-Shrinkage	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Elementary</i>							
Math pre-test score	-0.08* (0.01)	-0.20* (0.07)	-0.17* (0.05)	0.03* (0.01)	-0.07* (0.02)	-0.05* (0.02)	0.02* (0.01)
Reading pre-test score	-0.08* (0.01)	-0.14* (0.07)	-0.12* (0.06)	0.02 (0.01)	-0.06* (0.02)	-0.05* (0.02)	0.01* (0.01)
Free lunch eligible	0.05* (0.01)	0.34* (0.12)	0.29* (0.10)	-0.05* (0.02)	0.15* (0.03)	0.12* (0.03)	-0.03* (0.01)
Reduced-price lunch eligible	0.05 (0.04)	-0.80 (0.63)	-0.70 (0.52)	0.11 (0.12)	-0.10 (0.19)	-0.05 (0.15)	0.05 (0.05)
Limited English proficiency	0.05 (0.03)	-0.91* (0.21)	-0.79* (0.18)	0.12* (0.02)	-0.16* (0.07)	-0.11 (0.06)	0.05* (0.01)
Special education	0.22* (0.03)	0.22 (0.41)	0.18 (0.34)	-0.05 (0.07)	0.03 (0.11)	0.01 (0.09)	-0.01 (0.03)
Log student equivalents		-0.63* (0.07)	-0.54* (0.06)	0.09* (0.01)			
Student-equivalents /100					-0.09 (0.08)	-0.02 (0.06)	0.07* (0.02)
Number of observations (students)	7546	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Number of observations (teachers)	n.a.	297	297	297	297	297	297
<i>Middle School</i>							
Math pre-test score	-0.18* (0.01)	-0.64* (0.08)	-0.54* (0.07)	0.10* (0.01)	-0.01 (0.02)	0.01 (0.01)	0.02* (0.00)
Reading pre-test score	-0.15* (0.01)	-0.69* (0.08)	-0.58* (0.07)	0.11* (0.02)	-0.02 (0.02)	0.00 (0.02)	0.02* (0.01)
Free lunch eligible	0.08* (0.02)	0.67* (0.18)	0.57* (0.16)	-0.10* (0.03)	0.05 (0.04)	0.03 (0.03)	-0.02* (0.01)
Reduced-price lunch eligible	-0.02 (0.03)	-1.61 (0.83)	-1.38 (0.71)	0.23 (0.13)	-0.24 (0.15)	-0.15 (0.13)	0.09* (0.03)
Limited English proficiency	0.08* (0.03)	-0.26 (0.30)	-0.19 (0.25)	0.07 (0.05)	-0.03 (0.08)	-0.01 (0.07)	0.02 (0.01)

Characteristic	Log Variance of Individual Teacher Value-Added Estimates				Absolute Value of Individual Teacher Value-Added Estimates		
	Squared Residuals	Pre-Shrinkage	Post-Shrinkage	Difference	Pre-Shrinkage	Post-Shrinkage	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Special education	0.32* (0.03)	1.72* (0.35)	1.42* (0.29)	-0.30* (0.07)	-0.03 (0.07)	-0.08 (0.05)	-0.05 (0.03)
Log student equivalents		-0.66* (0.07)	-0.56* (0.06)	0.10* (0.01)			
Student-equivalents /100					0.01 (0.03)	0.04 (0.03)	0.02* (0.01)
Number of observations (students)	9930	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Number of observations (teachers)	n.a.	182	182	182	182	182	182

Source: Administrative data from OSSE and DCPS.

Note: Each group of cells displays the results from a separate regression. In column 1, the results in the first four rows are from a student-level regression of the squared residuals from the value-added model on the student characteristics. The remaining columns are teacher-level regressions. In columns 2 and 3, the outcome is the natural log of the value-added estimates' variance. Column 4 tests the statistical significance of the difference in the relationship between student characteristics and the natural log of the value-added estimates' variance. In columns 5 and 6, the outcome is the absolute value of the value-added estimates. Column 7 tests the statistical significance of the difference in the relationship between student characteristics and the absolute value of the value-added estimates. Robust standard errors are in parentheses.

*Significantly different from zero at the .05 level, two-tailed test.

DCPS = District of Columbia Public Schools; OSSE = Office of the State Superintendent of Education of the District of Columbia.

n.a. = not applicable.

Table 4. Effects of Student Characteristics on Distribution of Teacher Value-Added Estimates

Characteristic	Probability in Either the Bottom or Top Decile			Probability in Either the Bottom or Top Quintile		
	Pre-Shrinkage	Post-Shrinkage	Difference	Pre-Shrinkage	Post-Shrinkage	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Elementary School</i>						
Math pre-test score	-0.14* (0.04)	-0.14* (0.04)	0.00 (0.01)	-0.16* (0.05)	-0.15* (0.05)	0.01 (0.02)
Reading pre-test score	-0.12* (0.05)	-0.13* (0.05)	-0.01 (0.01)	-0.15* (0.05)	-0.14* (0.05)	0.01 (0.03)
Free lunch eligible	0.25* (0.07)	0.24* (0.07)	-0.01 (0.01)	0.42* (0.09)	0.36* (0.09)	-0.06 (0.04)
Reduced-price lunch eligible	-0.35 (0.39)	-0.23 (0.40)	0.12 (0.14)	-0.08 (0.46)	0.13 (0.47)	0.21 (0.21)
Limited English proficiency	-0.17 (0.13)	-0.15 (0.13)	0.02 (0.03)	-0.22 (0.18)	-0.22 (0.18)	0.00 (0.04)
Special education	-0.23 (0.24)	-0.12 (0.25)	0.11 (0.09)	0.20 (0.33)	-0.04 (0.32)	-0.24 (0.20)
Student-equivalents /100	-0.35* (0.16)	-0.32* (0.16)	0.04 (0.04)	-0.19 (0.25)	0.06 (0.25)	0.25 (0.14)
Number of teachers	297	297	297	297	297	297
<i>Middle School</i>						
Math pre-test score	-0.01 (0.06)	0.02 (0.06)	0.03 (0.05)	0.04 (0.07)	0.10 (0.07)	0.05 (0.03)
Reading pre-test score	-0.05 (0.07)	-0.02 (0.06)	0.04 (0.07)	0.05 (0.08)	0.09 (0.08)	0.03 (0.02)
Free lunch eligible	0.12 (0.14)	0.19 (0.13)	0.07 (0.12)	-0.05 (0.18)	-0.12 (0.18)	-0.07 (0.06)
Reduced-price lunch eligible	-0.51 (0.56)	-0.27 (0.55)	0.24 (0.32)	-0.69 (0.66)	-0.44 (0.67)	0.25 (0.22)
Limited English proficiency	-0.14 (0.19)	-0.14 (0.19)	0.00 (0.07)	-0.27 (0.28)	-0.25 (0.28)	0.02 (0.06)
Special	0.14	-0.13	-0.27	-0.28	-0.56	0.25

Characteristic	Probability in Either the Bottom or Top Decile			Probability in Either the Bottom or Top Quintile		
	Pre-Shrinkage (1)	Post-Shrinkage (2)	Difference (3)	Pre-Shrinkage (4)	Post-Shrinkage (5)	Difference (6)
education	(0.30)	(0.22)	(0.26)	(0.34)	(0.33)	(0.22)
Student-equivalents /100	-0.05 (0.11)	0.06 (0.11)	0.12 (0.10)	0.13 (0.14)	0.22 (0.14)	0.09 (0.06)
Number of teachers	182	182	182	182	182	182

Source: Administrative data from OSSE and DCPS.

Note: Each group of cells displays the results from a separate teacher-level regression. In columns 1 and 2, the outcome is the probability the teacher value-added estimate is below the 10th or above the 90th percentile of estimates. Column 3 tests the statistical significance of the difference in the relationship between student characteristics and the probability of being below the 10th or above the 90th percentile of estimates. In columns 4 and 5, the outcome is the probability the teacher value-added estimate is below the 20th or above the 80th percentile of estimates. Column 6 tests the statistical significance of the difference in the relationship between student characteristics and the probability the teacher value-added estimate is below the 20th or above the 80th percentile of estimates. Robust standard errors are in parentheses.

*Significantly different from zero at the .05 level, two-tailed test.

DCPS = District of Columbia Public Schools; OSSE = Office of the State Superintendent of Education of the District of Columbia.

About the Series

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information, contact Mariesa Herrmann, researcher, at mherrmann@mathematica-mpr.com, Elias Walsh, researcher, at ewalsh@mathematica-mpr.com, Eric Isenberg, associate director of research, at eisenberg@mathematica-mpr.com, or Alexandra Resch, researcher, at aresch@mathematica-mpr.com.

Authors' Note

We thank the Office of the State Superintendent of Education of the District of Columbia (OSSE) and the District of Columbia Public Schools (DCPS) for providing the data for this study. We are grateful to Duncan Chaplin and Cassandra Guarino for their helpful comments. Emma Kopa, assisted by Juha Sohlberg, Raúl Torres, and Maureen Higgins, provided excellent programming support. The paper was edited by John Kennedy. The text reflects the views and analyses of the authors alone and does not necessarily reflect views of Mathematica Policy Research, OSSE, or DCPS. All errors are the responsibility of the authors.

www.mathematica-mpr.com

**Improving public well-being by conducting high-quality,
objective research and surveys**

PRINCETON, NJ - ANN ARBOR, MI - CAMBRIDGE, MA - CHICAGO, IL - OAKLAND, CA - WASHINGTON, DC

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.