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

Mariesa Herrmann, Elias Walsh, and Eric Isenberg

ABSTRACT

It is common in the implementation of teacher accountability systems to use empirical Bayes shrinkage to adjust 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 article investigates how shrinkage affects the value-added estimates of teachers of hard-to-predict students. We found that teachers of students with low prior achievement and who receive free lunch tend to have less precise value-added estimates. However, in our sample, shrinkage had no statistically significant effect on the relative probability that teachers of hard-to-predict students received consequences.

1. Introduction

Due to the incentives provided by the federal Race to the Top program and Elementary Secondary Education Act (ESEA) waivers, districts and states rapidly implemented 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 incorporated 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 has been to use a value-added model (VAM) to estimate teachers’ contributions to student achievement on standardized assessments. Under the 2016 reauthorization of ESEA, the federal government gave states more control over their teacher evaluation systems. As states rethink their evaluation systems, there is an even greater need to understand how statistical models like VAMs contribute to assigning teachers to performance categories.

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 ability of the model to predict 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 with unusually high or low achievement growth can have a larger influence on the average differences between the actual and predicted performance of the students in the teacher’s class. Differences between students’ actual and predicted performance are related to 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 heteroscedasticity—variation across student groups in the difference between their actual and predicted achievement—has received relatively little attention in the value-added literature. However, Stacy et al. (2012) found that students with low socioeconomic status and low prior achievement had larger differences between actual and

ARTICLE HISTORY

Received January 2015
Accepted April 2016

KEYWORDS

Bayesian methods; Education; Measurement error
predicted achievement 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. The adjusted 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 have estimates in the extremes of the distribution. Similarly, if teachers of disadvantaged students have imprecise initial value-added estimates, shrinkage could reduce the probability that these teachers’ estimates are 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. Many 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.

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 article 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 heteroscedasticity, 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 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 article is to examine heteroscedasticity 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. 2015a), and other work on heteroscedasticity, 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.

2. Theory

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

\[ Y_{igt} = \lambda Y_{i,g-1} + \alpha' X_{igt} + \theta_t + \varepsilon_{igt}, \]  

(1)

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 \( X_{igt} \) denotes control variables for student demographic characteristics, \( \theta_t \) is a vector of teacher effects, and \( \varepsilon_{igt} \) is the student-level error term.

Under homoscedasticity, all student-level errors \( \varepsilon_{igt} \) 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 heteroscedastic errors. One possible cause of heteroscedasticity is measurement error in test scores. Typically, scores near the center of

---

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).

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.

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.
the distribution of a standardized test are more reliable than scores in the tails of the distribution (Resch and Isenberg 2014).6 Consequently, students who are likely to be at the top or the bottom of achievement distribution might tend to have large differences between their actual and predicted scores. Another reason for heteroscedasticity 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.

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 preshrinkage value-added estimate and the estimate for the average teacher. Let \( \hat{\theta}_t \) be the mean-centered preshrinkage point estimate for teacher \( t \) from the value-added regression model, and let the associated heteroscedasticity-robust variance estimate be \( \text{var}[\hat{\theta}_t] = \hat{\sigma}_t^2 \). 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:

\[
\hat{\theta}_{t|EB} \approx \left( \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_t^2} \right) \hat{\theta}_t,
\]

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 \( \hat{\sigma}_t^2 / (\hat{\sigma}_t^2 + \hat{\sigma}_t^2) \) must be less than 1, so the EB estimate always has a smaller absolute value than the initial estimate.

Although heteroscedasticity in student background characteristics does not affect the preshrinkage 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 preshrinkage estimate is given by the following:

\[
\frac{\partial \hat{\theta}_{t|EB}}{\partial \hat{\sigma}_t^2} = -\frac{\hat{\omega}_t^2}{\hat{\sigma}_t^2} \hat{\theta}_t,
\]

where \( \hat{\omega}_t \equiv \hat{\sigma}_t^2 / (\hat{\sigma}_t^2 + \hat{\sigma}_t^2) \), the weight on the preshrinkage 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 unshrunked 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 heteroscedasticity (or other characteristics of teachers or classrooms related to heteroscedasticity). 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 on 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 \( \hat{\sigma}_t^2 \) and \( \hat{\sigma}_t^2 \), the variance of a teacher’s EB estimate is approximately equal to

\[
\text{var} \left[ \hat{\theta}_{t|EB} \right] \approx \hat{\omega}_t \hat{\sigma}_t^2.
\]

Thus, the EB estimate has a smaller variance than that of the preshrinkage estimate, because \( \hat{\omega}_t \) is less than 1.

Heteroscedasticity affects \( \text{var} \left[ \hat{\theta}_{t|EB} \right] \) through the variance on the preshrinkage estimates. The change in the variance of a teacher’s EB estimate in response to a change in the variance of the preshrinkage estimate is proportional to the square of the weight on the point estimate in the EB calculation:

\[
\frac{\partial \text{var} \left[ \hat{\theta}_{t|EB} \right]}{\partial \hat{\sigma}_t^2} = \hat{\omega}_t^2.
\]

Because \( \hat{\omega}_t \) is less than one but greater than 0, shrinkage reduces but does not eliminate heteroscedasticity 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.

3. Value-Added Model and Data
We estimate a VAM that corresponds closely (though not exactly) with the model estimated for DCPS for teacher evaluations in the 2011–2012 school year (described in Isenberg and Hock 2012). This model has many features that are common in VAMs estimated in other states or districts. Six of nine

---

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, Leatherman, and Parsons (2012) developed a method for accounting for measurement error that varies across the achievement distribution.

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.

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.
teacher-level VAMs for which we could obtain documentation use teacher fixed effects and apply a post-hoc adjustment to apply empirical Bayes’ shrinkage to the fixed effect estimates, the method we use for our analysis. The other three VAMs bring about shrinkage as part of the regression model by using teacher random effects. The popularity of the two-step approach may be due in part to the desire to avoid the assumption of a random effects model that teacher effectiveness is uncorrelated with student characteristics. Using simulated data, Guarino, Reckase, and Wooldridge (2015b) suggest that teacher fixed effects models are more robust to circumstances in which there is systematic sorting of students to teachers—that is, to the type of sorting that would violate this assumption of a random effects model.

We estimate the VAM separately for teachers of elementary (grades 4 and 5) and middle school (grades 6 through 8) students:

\[ Y_{tig} = \lambda_g Y_{ig(-1)} + \omega_g Z_{ig(-1)} + \alpha'X_i + \delta_g + \eta'T_{ti} + \epsilon_{tig}, \]  

where \( Y_{tig} \) is the posttest score for student \( i \) in grade \( g \) linked to teacher \( t \) and \( Y_{ig(-1)} \) is the same-subject pretest for student \( i \) in grade \( g-1 \) during the previous year. The variable \( Z_{ig(-1)} \) denotes the pretest in the opposite subject. Thus, when estimating teacher effectiveness in math, \( Y_g \) and \( Y_{ig(-1)} \) represent math tests, with \( Z_{ig(-1)} \) representing reading tests and vice versa. The pretest scores capture prior inputs into student achievement, and the associated coefficients \( \lambda_g \) and \( \omega_g \) 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 \( \delta_g \) includes grade indicators.

The vector \( T_{ti} \) 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. The vector \( \eta \) includes one coefficient for each teacher. Finally, \( \epsilon_{tig} \) is the random error term.

Measurement error in the pretest scores will attenuate the estimated relationship between the pre- and posttest scores. To avoid this bias, the VAM in Equation (6) is estimated in two regression steps to account for both measurement error in the pretests and clustering of standard errors at the student level, which cannot be implemented simultaneously. In the first step, we estimate Equation (6) adjusting for measurement error in the pretests using an errors-in-variables correction (eivreg in Stata) that relies on published information on the test-retest reliability of the District of Columbia Comprehensive Assessment System (DC CAS) (Buonaccorsi 2010). We then used the measurement-error corrected values of the pretest coefficients to calculate adjusted posttest scores by subtracting the predicted effects of the pretest scores from the posttest scores:

\[ \hat{G}_{tig} = Y_{tig} - \hat{\lambda}_g Y_{ig(-1)} - \hat{\omega}_g Z_{ig(-1)}. \]  

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 heteroscedasticity.

In the second step, we then estimate teacher effects by regressing the adjusted posttest scores from the first step on student background characteristics, grade indicators, and teacher indicators, clustering standard errors by student using a cluster-robust sandwich variance estimator (Liang and Zeger 1986; Arellano 1987):

\[ \hat{G}_{tig} = \alpha'X_i + \delta_g + \eta'T_{ti} + \epsilon_{tig}. \]  

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. We then remove any teachers with fewer than 15 students, and re-center the value-added estimates to have a mean of 0. Finally, we 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).

We use data from DCPS and the Office of the State Superintendent of Education of the District of Columbia (OSSE). The data contain information on students’ test scores and demographic characteristics and enable students to be linked to their teachers. In DCPS, students in grades 3 through 8 and 10 were tested in math and reading using the 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 posttest 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.

Teachers vary in the characteristics of students that they teach. Table 1 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. Excluding teachers who are matched to fewer than 15 students follows the districts’ implementation of its teacher accountability system.
Table 1. Teacher-level summary statistics.

| Variables                          | Mean   | Standard deviation | Minimum   | Maximum   |
|------------------------------------|--------|--------------------|-----------|-----------|
| **Elementary School**              |        |                    |           |           |
| Math posttest score                | 0.02   | 0.57               | 1.69      | 1.26      |
| Math pretest score                 | 0.02   | 0.55               | 1.34      | 1.36      |
| Reading pretest 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     |
| Preshrinkage value-added estimates | 0.01   | 0.30               | −0.90     | 0.96      |
| Estimated variance of preshrinkage 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    |                    |           |           |
| **Middle School**                  |        |                    |           |           |
| Math posttest score                | −0.04  | 0.49               | −1.27     | 1.17      |
| Math pretest score                 | −0.05  | 0.50               | −1.12     | 1.16      |
| Reading pretest 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    |
| Preshrinkage value-added estimates | −0.01  | 0.20               | −0.58     | 0.63      |
| Estimated variance of preshrinkage 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.

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 1 shows that there is wide variation in students’ test scores and classroom composition across teachers. For example, in math, the average pretest 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 1 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.

Postshrinkage value-added estimates and variances are smaller than the preshrinkage estimates. Table 1 presents the preshrinkage and shrinkage value-added estimates and their estimated variances. For elementary school teachers, the preshrinkage estimates range from −0.90 to 0.96 and the estimated preshrinkage variances range from 0 to 0.07. Because the shrinkage procedure assigns the preshrinkage 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 preshrinkage estimates and the shrinkage variances are smaller than the preshrinkage variances. For elementary school teachers, the postshrinkage estimates range from −0.77 to 0.71 and the estimated postshrinkage 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.

### 4. Empirical Approach

To describe heteroscedasticity 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:

$$\hat{\epsilon}_{tig}^2 = \alpha + \beta X_i + \nu_{tig}.$$  \hfill (9)
The outcome $\varepsilon_{ig}^2$ 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 $X_t$ denotes student characteristics (that is, average math and reading pretest scores, eligibility for free or reduced-price lunches, limited English proficiency, and special education status), and $\nu_{igt}$ is an error term.

The regression in Equation (9) is intended to provide a parsimonious description of heteroscedasticity 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. 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. As a sensitivity analysis, we also estimate Equation (9) using the natural log of the squared residual as the dependent variable.

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 postshrinkage value-added estimates. We estimate the following teacher-level bivariate regression model, again using separate regressions for each student characteristic and grade span:

$$W_t = \alpha + \beta X_t + \epsilon_t,$$

(10)

where $W_t$ represents a measure of the precision or magnitude of the pre- or postshrinkage value-added estimate for teacher $t$. Specifically, for both pre- and postshrinkage 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. We examine the absolute value rather than the original value of the estimate because value-added estimates have mean zero (for the average teacher), so the further from zero, the more extreme the estimate of teacher effectiveness in a good (positive) or bad (negative) direction. When $W_t$ represents an indicator for being above or below a cut-point, we present results based on a linear probability model, but as a sensitivity analysis we also estimate a logit model. The vector $X_t$ represents the average characteristics of the teacher’s students from Table 1, 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.

### 5. Results

Table 2 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 (Equation (9)) on student characteristics and provides evidence that some groups of students have significantly larger residuals than others. This is consistent with heteroscedasticity 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 pretest scores.\(^\text{13}\)

The same characteristics are also significantly associated with larger residuals in middle school, as is limited English proficiency. When we substituted the natural log of the squared residual as the dependent variable, we obtained results with very similar results: for each variable, the direction (positive or negative) was identical to the results shown in Table 2, and the same variables were statistically significant (not shown). 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 2, we provide the results from teacher-level regressions, which separately regress the natural log of teachers’ preshrinkage variances on each of the characteristics of teachers’ students and the natural log of the teachers’ number of student-equivalents (Equation (10)).\(^\text{14}\) For elementary school teachers, a one-standard-deviation decrease in students’ average pretest scores in math is significantly associated with a 20% increase in the teacher’s preshrinkage variance. Lower pretest scores in reading and free lunch eligibility are also significantly associated with higher preshrinkage variances in elementary school. Interestingly, limited English proficiency is significantly correlated with lower preshrinkage variances in elementary school, and there is no significant relationship between the preshrinkage

---

\(^\text{13}\) To determine whether students with very high or very low pretest 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 pretest scores, an indicator for having a below-mean pretest score, and an interaction between the absolute value of the pretest score and having a below-mean pretest score (results not shown). Although having a pretest score at either extreme of the distribution is associated with larger residuals, residuals are significantly higher for students at the bottom of the pretest distribution than for those at the top.

\(^\text{14}\) 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: $\ln(\hat{V}_t) = \ln(\hat{\sigma}_t^2) - \ln(N_t) = \ln(\alpha + \beta X_t + \epsilon_t) - \ln(N_t)$, where $\hat{V}_t$ is the estimated variance and $N_t$ is the number of student equivalents.
Table 2. Effects of student characteristics on teacher value-added estimates.

| Characteristic                     | Squared residuals (1) | Log variance of individual teacher value-added estimates | Absolute value of individual teacher value-added estimates |
|-----------------------------------|-----------------------|--------------------------------------------------------|----------------------------------------------------------|
|                                   | Pre shrinkage (2)     | Post shrinkage (3) | Difference (4)                                          | Pre shrinkage (5)     | Post shrinkage (6) | Difference (7) |
| Elementary                        |                       |                                                          |                                                          |                                                          |
| Math pretest score                | −0.08*                | −0.20*                                                   | −0.17*                                                   | 0.03*                                                      | −0.07*            | −0.05*          | 0.02*          |
| (0.01)                            | (0.07)                | (0.05)                                                  | (0.01)                                                  | (0.02)                                                     | (0.02)            | (0.01)          |               |
| Reading pretest score             | −0.08*                | −0.14*                                                   | −0.12*                                                   | 0.02                                                      | −0.06*            | −0.05*          | 0.01*          |
| (0.01)                            | (0.07)                | (0.06)                                                  | (0.01)                                                  | (0.02)                                                     | (0.02)            | (0.01)          |               |
| Free lunch eligible               | 0.05*                 | 0.34*                                                    | 0.29*                                                   | −0.05*                                                     | 0.15*             | 0.12*           | −0.03*         |
| (0.01)                            | (0.12)                | (0.10)                                                  | (0.02)                                                  | (0.03)                                                     | (0.03)            | (0.01)          |               |
| Reduced-price lunch eligible      | 0.05                  | −0.80                                                   | −0.70                                                   | 0.11                                                      | −0.10             | −0.05           | 0.05           |
| (0.04)                            | (0.63)                | (0.52)                                                  | (0.12)                                                  | (0.19)                                                     | (0.15)            | (0.05)          |               |
| Limited English proficiency       | 0.05                  | −0.91*                                                   | −0.79*                                                   | 0.12*                                                     | 0.16*             | −0.11           | 0.05*          |
| (0.03)                            | (0.21)                | (0.18)                                                  | (0.02)                                                  | (0.07)                                                     | (0.06)            | (0.01)          |               |
| Special education                 | 0.22*                 | 0.22                                                    | 0.18                                                    | −0.05                                                     | 0.03              | 0.01            | −0.01          |
| (0.03)                            | (0.41)                | (0.34)                                                  | (0.07)                                                  | (0.11)                                                     | (0.09)            | (0.03)          |               |
| Log student equivalents           | −0.63*                | −0.54*                                                   | 0.09*                                                   |                                                         | −0.09             | −0.02           | 0.07*          |
| (0.07)                            | (0.06)                | (0.01)                                                  |                                                          | (0.08)                                                     | (0.06)            | (0.02)          |               |
| Student-equivalents / 100         |                       |                                                          |                                                          |                                                          |                   |                 |               |
| Number of observations (students) | 7546                  | n.a.                                                    | n.a.                                                    | n.a.                                                       | 297               | 297             | 297            |
| Number of Observations (teachers) | n.a.                  | 297                                                     | 297                                                     | 297                                                        |                   |                 |               |
| Middle School                     |                       |                                                          |                                                          |                                                          |                   |                 |               |
| Math pretest score                | −0.18*                | −0.64*                                                   | −0.54*                                                   | 0.10*                                                      | −0.01             | 0.01            | 0.02*          |
| (0.01)                            | (0.08)                | (0.07)                                                  | (0.01)                                                  | (0.02)                                                     | (0.02)            | (0.01)          |               |
| Reading pretest score             | −0.15*                | −0.69*                                                   | −0.58*                                                   | 0.11*                                                      | −0.02             | 0.00            | 0.02*          |
| (0.01)                            | (0.08)                | (0.07)                                                  | (0.02)                                                  | (0.02)                                                     | (0.02)            | (0.01)          |               |
| Free lunch eligible               | 0.08*                 | 0.67*                                                    | 0.57*                                                   | −0.10*                                                     | 0.05              | 0.03            | −0.02*         |
| (0.02)                            | (0.18)                | (0.16)                                                  | (0.03)                                                  | (0.04)                                                     | (0.03)            | (0.01)          |               |
| Reduced-price lunch eligible      | −0.02                 | −1.61                                                   | −1.38                                                   | 0.23                                                      | −0.24             | −0.15           | 0.09*          |
| (0.03)                            | (0.83)                | (0.71)                                                  | (0.13)                                                  | (0.15)                                                     | (0.13)            | (0.03)          |               |
| Limited English proficiency       | 0.08*                 | −0.26                                                   | −0.19                                                   | 0.07                                                      | −0.03             | −0.01           | 0.02           |
| (0.03)                            | (0.30)                | (0.25)                                                  | (0.05)                                                  | (0.08)                                                     | (0.07)            | (0.01)          |               |
| Special education                 | 0.32*                 | 1.72*                                                    | 1.42*                                                   | −0.30*                                                     | −0.03             | −0.08           | −0.05          |
| (0.03)                            | (0.35)                | (0.29)                                                  | (0.07)                                                  | (0.07)                                                     | (0.05)            | (0.03)          |               |
| Log student equivalents           | −0.66*                | −0.56*                                                   | 0.10*                                                   |                                                         | 0.01              | 0.04            | 0.02*          |
| (0.07)                            | (0.06)                | (0.01)                                                  |                                                          | (0.03)                                                     | (0.03)            | (0.01)          |               |
| Student-equivalents / 100         |                       |                                                          |                                                          |                                                          |                   |                 |               |
| Number of observations (students) | 9930                  | n.a.                                                    | n.a.                                                    | n.a.                                                       | 182               | 182             | 182            |
| Number of observations (teachers) | n.a.                  | 182                                                     | 182                                                     | 182                                                        |                   |                 | 182            |

Source: Administrative data from OSSE and DCPS.

NOTE: Each result is from a separate regression except for those for eligibility for free or reduced-price lunch, which were entered simultaneously into a 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 0.05 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.

Variances and special education status. Having a larger number of students is associated with lower preshrinkage variances. The same general patterns hold in middle school. Lower pretest scores in math and reading, free lunch eligibility, and special education status are significantly associated with higher preshrinkage 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% increase in a teacher’s preshrinkage variance in the middle school grades. Column 2 of Table 2 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 postshrinkage 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 2. In column 3, we

---

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.
examine the relationships among teachers’ postshrinkage variances, the characteristics of their students, and teachers’ number of student-equivalents. Compared with the preshrinkage variances, the postshrinkage variances have the same number of statistically significant relationships with student characteristics, although the size of these correlations is significantly smaller for the postshrinkage variances. For example, for elementary school students, a one standard deviation decrease in students’ average pretest scores in math is significantly associated with a 17% increase in the teacher’s postshrinkage variance, compared with a 20% increase in the preshrinkage variance, a statistically significant difference. Statistical tests of the differences between the coefficients for preshrinkage and postshrinkage 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 2 presents results from teacher-level regressions of the absolute value of teachers’ preshrinkage value-added estimates on student characteristics and teachers’ number of student-equivalents. In elementary school, the preshrinkage value-added estimates for teachers of students with lower pretest scores and teachers with more students eligible for free lunch are significantly further from (and with limited English proficiency closer to) the overall teacher mean relative to the preshrinkage value-added estimates of other teachers. In middle school, student characteristics do not have a significant relationship with the absolute value of the preshrinkage 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’ postshrinkage 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 shrinkage estimates, but these relationships are weaker than their relationships with the preshrinkage 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. 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. 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 postshrinkage 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 3. In column 1, using a linear probability model, we regress an indicator variable for the preshrinkage estimate falling in either the top or bottom deciles of the preshrinkage value-added distribution on each student characteristic and on teachers’ number of student-equivalents (Equation 10). For elementary school, we find that having students with low pretest scores, students eligible for free lunch, or smaller numbers of students increases the probability that teachers’ preshrinkage 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 postshrinkage estimate is either in the top or bottom deciles of the postshrinkage value-added distribution and column 3 tests whether the difference between the pre- and postshrinkage relationships is statistically significant. Of 57 elementary school teachers in the top or bottom decile before applying shrinkage, there were 2 who moved in out of the extremes of the distribution after applying shrinkage. For middle school teachers, 5 of 32 teachers did likewise. 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. Comparing movement of teachers in and out of the top and bottom quintiles, 7 of 119 elementary school teachers switched places, as did 3 of 73 middle school teachers. However, shrinkage does not significantly affect the probability that teachers of hard-to-predict students

---

56 The absolute value is measured relative to the overall teacher mean, rather than the mean conditional on the characteristics of teachers’ students. The preshrinkage 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 2 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.

57 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.

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

59 In a typical teacher evaluation system, like the system that was in place in DCPS, this does not force a certain number of teachers to have positive consequences and others to have negative consequences because value-added estimates are used in conjunction with rigorous teacher observations and other measures to create a final evaluation score for each teacher.
are classified in either the top or bottom quintiles of the value-added distribution. 20

6. Conclusion

Recent work has raised concerns about the effect of heteroscedasticity on estimates of value added because heteroscedasticity can reduce the precision of value-added estimates for teachers of particular types of students (e.g., Stacy et al. 2012). Because shrinkage adjusts value-added estimates by their level of precision, it might be viewed as a method to mitigate heteroscedasticity. This article 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. We find three main results. First, consistent with heteroscedasticity, we find that particular student characteristics associated with disadvantage—lower pretest scores and eligibility for free lunch—are correlated with the precision of a teacher’s estimate. Second, 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. Third, in these data, shrinking has no statistically significant impact on the relative probability of exceeding a threshold for teachers of hard-to-predict students.

The first two results are predicted by theory, while the third result is puzzling. One piece of this puzzle may be that we assume that the accountability thresholds are rescaled along with the estimates. So, while shrinkage moves all teachers toward the middle of the distribution, the thresholds also adjust inward as shrinkage reduces the variance of teacher estimates. One must use caution in generalizing the result that, essentially, the accountability thresholds move in as quickly as teachers of

---

20 As a sensitivity analysis, we estimated Equation (10) using a logit model in place of a linear probability model. The direction and statistical significance of coefficients reported in columns 1, 2, 4, and 5 of Table 3 are nearly identical (not shown). The only difference is for the coefficient for the number of student-equivalents. These coefficients are just under the threshold for statistical significance in the elementary grade span for the top/bottom decile specification using a linear probability model. By contrast, they are not statistically significant when using a logit model.
disadvantaged students are pulled toward the middle. In this investigation, we have focused on data from a single year and district. An analysis using more years and districts would reveal whether this result is an anomaly or part of a broader pattern. In the meantime, given the conceptual appeal of shrinkage, the relative simplicity of implementation as a postestimation step, and the importance of this procedure in protecting teachers with relatively few students from being placed in the extremes of the distribution by chance, we recommend its use as part of any value-added model used for teacher accountability.

Acknowledgments

We thank the Office of the State Superintendent of Education of the District of Columbia and the District of Columbia Public Schools for providing the data for this study. We are grateful to Duncan Chaplin, Alexandra Resch, Cassandra Guarino, and two anonymous referees for their helpful comments. Emma Kopa, assisted by Juha Sohlberg, Raúl Torres, and Maureen Higgins, provided excellent programming support.

References

Aaronson, D., Barrow, L., and Sander, W. (2007), “Teachers and Student Achievement in the Chicago Public High Schools,” *Journal of Labor Economics*, 25, 95–135. [1]

American Institutes for Research (2013), “Florida Comprehensive Assessment Test (FCAT) 2.0 Value-Added Model: Technical Report 2012–13,” Washington, DC: American Institutes for Research. [4]

American Institutes for Research (2014), “Baltimore City Public Schools Value-Added Model Technical Report,” Washington, DC: American Institutes for Research. [4]

Arellano, M. (1987), “Computing Robust Standard Errors for Within-Groups Estimators,” *Oxford Bulletin of Economics and Statistics*, 49, 431–434. [4]

Buonaccorsi, J. P. (2010), *Measurement Error: Models, Methods, and Applications*, Boca Raton, FL: Chapman & Hall/ CRC. [4]

Chetty, R., Friedman, J., and Rockoff, J. (2014), “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” *American Economic Review*, 104, 2593–2632. [1]

Guarino, C. M., Maxfield, M., Reckase, M. D., Thompson, P., and Wooldridge, J. M. (2015), “An Evaluation of Empirical Bayes’ Estimation of Value-Added Teacher Performance Measures,” *Journal of Educational and Behavioral Statistics*, 40, 190–222. [2]

Guarino, C. M., Reckase, M. D., and Wooldridge, J. M. (2015), “Can Value-Added Measures of Teacher Performance be Trusted?” *Education Finance and Policy*, 10, 117–156. [1,4]

Hock, H., and Isenberg, E. (2012), *Methods for Accounting for Co-Teaching in Value-Added Models*, Working Paper No. 6, Washington, DC: Mathematica Policy Research. [4]

Isenberg, E., and Hock, H. (2012), “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. [3]

Isenberg, E., and Walsh, E. (2014), “Measuring Teacher Value Added in DC, 2013–2014 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. [4]

Isenberg, E., and Walsh, E. (2015), “Accounting for Co-Teaching: A Guide for Policymakers and Developers of Value-Added Models,” *Journal of Research on Educational Effectiveness*, 8, 112–119. [4]

Kane, T. J., and Staiger, D. O. (2002), “The Promise and Pitfalls of Using Imprecise School Accountability Measures,” *Journal of Economic Perspectives*, 16, 91–114. [2]

Kane, T. J., and Staiger, D. O. (2008), Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation, Working Paper No. 14607, Cambridge, MA: National Bureau of Economic Research. [1]

Kane, T., McCaffrey, D., Miller, T., and Staiger, D. (2013), *Have we Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*, Seattle, WA: Bill & Melinda Gates Foundation. [1]

Koedel, C., Leatherman, R., and Parsons, E. (2012), *Test Measurement Error and Inference from Value-Added Models*, unpublished manuscript, Columbia, MO: University of Missouri. [3]

Liang, K.-Y., and Zeger, S. L. (1986), “Longitudinal Data Analysis Using Generalized Linear Models,” *Biometrika*, 73, 13–22. [4]

Louisiana’s Value-Added Assessment Model as Specified in Act 54 (2013), “A Report to the Board of Elementary and Secondary Education, September 2013. Available at [http://www.louisianabelieves.com/docs/teaching/2012-2013-value-added-report.pdf?sfvrsn=8](http://www.louisianabelieves.com/docs/teaching/2012-2013-value-added-report.pdf?sfvrsn=8). Accessed March 31, 2016. [4]

Mihaly, K., McCaffrey, D. F., Staiger, D. O., and Lockwood, J. R. (2013), A Composite Estimator of Effective Teaching, Seattle, WA: Bill & Melinda Gates Foundation. [1]

Morris, C. N. (1983), “Parametric Empirical Bayes Inference: Theory and Applications,” *Journal of the American Statistical Association*, 78, 47–55. [3]

Resch, A., and Deutsch, J. (2015), “Measuring School and Teacher Value Added in Charleston County School District, 2014–2015 School Year,” Final report submitted to the Charleston County School District. Washington, DC: Mathematica Policy Research. [4]

Resch, A., and Isenberg, E. (2014), “How Do Test Scores at the Floor and Ceiling Affect Value-Added Estimates?” Working paper 33. Washington, DC: Mathematica Policy Research. [3]

Rothstein, J. (2010), “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement,” *Quarterly Journal of Economics*, 125, 175–214. [1]

Rotz, D., Johnson, M., and Gill, B. (2014), “Value-Added Models for the Pittsburgh Public Schools, 2012-13 School Year,” Cambridge, MA: Mathematica Policy Research. [4]

Stacy, B., Guarino, C., Reckase, M., and Wooldridge, J. (2012), “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. [1,2,6,9]

Tate, R. L. (2004), “A Cautionary Note on Shrinkage Estimates of School and Teacher Effects,” *Florida Journal of Educational Research*, 42, 1–21. [2]

Value-Added Research Center (2010), *NYC Teacher Data Initiative: Technical Report on the NYC Value-Added Model 2010*, Madison, WI: Value-Added Research Center at Wisconsin Center for Education Research, University of Wisconsin-Madison. [4]

——— (2014), *Hillsborough County Public Schools Value-Added Project: Year 3 Technical Report HCPSS Value-Added Models 2013*, Madison, WI: Value-Added Research Center at Wisconsin Center for Education Research, University of Wisconsin-Madison. [4]

Walsh, E., Liu, A. Y., and Dotter, D. (2015), “Measuring Teacher and School Value Added in Oklahoma, 2013–2014 School Year,” final report submitted to the Office of Educator Effectiveness at the Oklahoma S.0.tate Department of Education. Washington, DC: Mathematica Policy Research. [4]