

# The Role of Teacher Private Information in Predicting Student Test Scores\*

Esteban M. Aucejo<sup>†</sup>                      Ilkem Gok Karci<sup>‡</sup>  
*Arizona State University*              *Arizona State University*

Paola Ugalde Araya<sup>§</sup>  
*Louisiana State University*

February 11, 2026

## Abstract

This paper investigates whether incorporating teacher private information (TPI) on students' skills can improve the validity of value-added (VA) models in education. Using administrative data on teacher–student assignments, we show that a coarse measure of TPI significantly predicts students' future academic outcomes, even after controlling for a rich set of student characteristics. For example, our measure of TPI has as much predictive power for future math achievement as the combined effect of student gender, race, special education, economic disadvantage and English learner statuses (ELS). Notably, teachers' private information of students' level of proficiency diverges more from test-based score levels for Black students, economically disadvantaged, English learners, and students in special education, suggesting that standardized tests may fail more often to fully capture the skills of these groups. We also find that TPI correlates with VA estimates when excluded from the model, indicating potential selection on observables. Incorporating TPI meaningfully alters teacher rankings based on VA, highlighting its role in improving both the accuracy and fairness of evaluation. Taken together, these findings underscore the value of leveraging teachers' private information to enhance the robustness and equity of VA models and to more effectively identify persistently low-performing teachers.

**Keywords:** value-added, teacher, private information

**JEL Classification Codes:** I2, I20, I21

---

\*The views expressed are those of the authors and do not necessarily reflect the positions or policies of the North Carolina Department of Public Instruction.

<sup>†</sup>Dept of Economics, Arizona State University, CEP, & NBER, United States, [Esteban.Aucejo@asu.edu](mailto:Esteban.Aucejo@asu.edu)

<sup>‡</sup>Dept of Economics, Arizona State University, United States, [igokkarc@asu.edu](mailto:igokkarc@asu.edu)

<sup>§</sup>Dept of Economics, Louisiana State University, United States, [pugaldea@lsu.edu](mailto:pugaldea@lsu.edu)

# 1 Introduction

Evaluating teacher performance has long been a central issue in education research and policy. Among various methodologies, value-added (VA) models, which estimate a teacher’s contribution to student achievement by controlling for prior test scores and other observable characteristics, have attracted significant attention. Proponents argue that VA estimates can guide policy decisions, inform teacher retention and compensation strategies, and ultimately enhance student outcomes (Hanushek, 2009; Chetty et al., 2014b). Yet, the validity of VA estimates remains contentious, as skeptics point to potential biases and instability that arise from relying heavily on test scores and limited observable data (Koedel and Rockoff, 2015; Rothstein, 2017). For example, a key challenge in measuring teacher effectiveness with VA models is the presence of measurement error in prior student skills. Standardized assessments may capture only a fraction of a student’s knowledge and skills, leaving unmeasured components (whether due to temporary shocks on test day or more persistent, subtle aspects of ability) unaccounted for. These gaps can create distortions in VA estimates, raising concerns that high-stakes decisions are based, at least in part, on noisy or not totally valid measures of teacher effectiveness.

This paper highlights an overlooked but potentially important source of improvement in VA estimation: teachers’ private information about their students. Teachers, through daily interactions and classroom assessments, often possess insights into a student’s ability and understanding that standardized tests fail to detect. This private information may help refine predictions of future student performance, thereby mitigating the measurement error in the lagged test scores and problems of omitted variable bias. Our study incorporates a proxy for teacher-specific private knowledge, constructed from the discrepancy between a teacher’s *ex-ante* expectations about students’ test performance and the student’s realized score. Our findings demonstrate that adding even coarse measures of teacher private information can impact VA estimates in a way that could influence policy recommendations.

A range of solutions has been proposed to mitigate issues associated with measuring

students' prior skills. Some recommend using multiple test administrations per subject to better isolate true achievement (Boyd et al., 2013), while others advocate pooling information across years and outcomes (Kane and Staiger, 2002). Nevertheless, these methods can be data-intensive, complex, or may fail to solve the core challenge of capturing students' latent skills. In general, the literature has acknowledged the problem of measurement error (or omitted variable bias) in prior skills but has not focused on leveraging teachers' private insights as a remedy.

Using data from North Carolina public schools, the aim of this paper is threefold. First, we study whether teachers possess important private information about their students' academic skills that traditional measures of student backgrounds cannot fully capture. Second, we analyze the type of students for whom this private information could be more relevant. Finally, we assess whether incorporating teachers' private knowledge into value-added (VA) models can impact VA estimates in a way that also alters the composition of teachers at the bottom of the VA distribution—those whom previous research, such as Chetty et al. (2014b) and Hanushek (2009), suggests might be replaced to improve student outcomes.

We find that our proxy measure of teacher private information (TPI) significantly predicts students' next academic year performance. Notably, the predictive power of TPI remains strong and statistically significant even after accounting for an extensive set of control variables. In fact, TPI alone is nearly as predictive as gender, race, special education status, English learner status, and the economic disadvantage status of the students, all combined.

Second, we demonstrate that teachers' private information is more relevant for evaluating students whose abilities may not be accurately reflected by standardized test scores. Previous research indicates that standardized tests frequently underestimate the skills of lower-income students and students of color (Rothstein, 2004; Geiser and Santelices, 2007). Consistent with these findings, our analysis reveals that teacher assessments tend to diverge from standardized test scores for several key student subgroups. Specifically, discrepancies are more pronounced among Black students, those who are economically disadvantaged, En-

glish learners, or have special education needs, and those with lower prior test scores, with differences observed in both math and reading assessments.

Finally, we show a statistically significant correlation between TPI and value-added (VA) estimates when TPI is omitted from the VA model. This suggests the presence of selection on observables, which is consistent with our evidence of forecast bias. Including TPI in VA estimation materially alters the identification of low-performing teachers. Specifically, when comparing models with and without TPI, only 67.4% of math teachers and 63% of reading teachers initially ranked in the bottom 5% of the VA distribution remain in that category. The fact that a coarse measure of TPI both strongly predicts future test scores and influences VA estimates highlights the need to explore more refined proxies for teacher private information to better capture teacher effectiveness.

In summary, this study highlights the importance of incorporating measures of TPI into VA models. By leveraging teachers' unique, context-specific insights, our findings suggest that improvements in teacher evaluation systems are both achievable and policy-relevant. Recognizing the value of teachers' private information, in turn, may enhance the validity and fairness of high-stakes personnel decisions in education.

The rest of the paper is structured as follows. Section 2 describes the data and the North Carolina context. Section 3 provides evidence for the role of teachers' private information in predicting students' test scores. Section 4 analyzes the type of students to whom this private information is more relevant. Section 5 examines the extent to which TPI impacts value-added estimates. Section 6 concludes.

## 2 Data

Our analysis draws on administrative data from the North Carolina Education Research Data Center (NCRDC) spanning 2007 to 2011, which includes teacher predictions of students' end-of-grade (EOG) performance alongside actual test scores. Because the data iden-

tify proctors—who typically serve as the students’ main teacher—we focus on grades 4 and 5 (usually self-contained classes) in which the same proctor administered both math and reading exams.<sup>1</sup>

Panel A of Appendix Table A1 shows 762,108 student-year observations in our dataset. Of these students, 50% are female, 59% are White, 21% are Black, and 9% are Hispanic. Approximately 29% qualify as economically disadvantaged (EDS), 5% have English learner status (ELS), and 3% are enrolled in special education. We also have access to standardized end-of-grade math and reading scores by grade and year. Panel B shows data at the teacher-year level, comprising 37,336 observations, where 90% of teachers are female, 87% are White, and 11% are Black.

Teachers assign each student an expected academic level (from 1 to 4) on the day of the End-of-Grade (EOG) exam (i.e., before knowing the actual performance of the students).<sup>2</sup> This scale mimics the student’s test score performance level, making any discrepancy between the predicted and actual level a proxy for teacher private information (TPI) that is not captured by the EOG exam or, conversely, teacher error.<sup>3</sup> Ultimately, the extent to which TPI predicts student performance in subsequent years will determine whether these discrepancies reflect meaningful private information or mere teacher prediction error.

We operationalize TPI in two ways. The first is a discrete measure ranging from -3 (the teacher underestimates the student by three levels) to +3 (the teacher overestimates by three levels) that represents the difference in levels between the teacher prediction and the test score level performance. The second is a continuous measure derived by subtracting a student’s actual standardized EOG score from the median standardized score associated

---

<sup>1</sup>Under this restriction, 99.85% of observations have the same proctor; the remaining 0.32% corresponds to students who only took one exam. Grades 1 and 2 are excluded because they do not take EOG exams, and we exclude third-grade observations because there are no teacher predictions from the previous year.

<sup>2</sup>More specifically, we rely on the variable teacher judgment, where 1: Teacher judges that this student does not have sufficient mastery, 2: Student demonstrates inconsistent mastery of knowledge, 3: Student consistently demonstrates mastery, 4: Student consistently performs in a superior manner.

<sup>3</sup>Achievement levels in the EOG exams are as follows: 1: Insufficient mastery, 2: Inconsistent mastery, 3: Consistent mastery, 4: Superior performance. These levels are a function of student scores in the EOG exam (e.g., students scoring below a certain value in the exam belong to level 1, etc).

with the teacher’s predicted level.<sup>4</sup> This continuous metric provides more granular insight into teacher judgment within each achievement level. As shown in columns (1) and (2) of Table 1, nearly 40% of teacher predictions deviate from the “0” category, suggesting that teacher assessments frequently differ from students’ actual test performance. Furthermore, the discrepancies are distributed relatively symmetrically around zero, suggesting no large systematic biases in one particular direction. The prevalence of these discrepancies suggests that, rather than systematic misjudgment, teachers may be capturing relevant aspects of student ability that standardized tests fail to measure. This interpretation implies that incorporating TPI into models of student achievement could provide a more comprehensive understanding of student learning while mitigating the potential drawbacks of mainly relying on test-based evaluations. Columns (3) to (6) show that in many instances, the magnitude of the (proxy) discrepancies could be very large when measured in test scores standard deviation units (i.e., using our continuous measure of TPI), underscoring the potential limitations of relying solely on standardized assessments. To further show that discrepancies in predictions are not mainly driven by the marginal students in each level (i.e., those right below or above the threshold of a given achievement level), we reconstructed Table 1 but this time dropping the students within half a standard deviation of the cut-off level values. In this regard, Table A2 shows that even after making these adjustments, a large share of teacher predictions differed from students’ actual performance.

Finally, Tables A3 and A4 present the average student level reported by teachers, conditional on students’ actual test score performance. For example, Table A3 shows that among students who scored at level one, the average level reported by their teachers was 1.94 in math and 2.03 in reading. The data also indicate that alignment between teacher predictions and actual test scores is more frequent at level three, where the mean teacher-reported level for students scoring at this level was 3.05 in math and 3 in reading. Table A4 examines the magnitude of discrepancies using our continuous measure, conditional on students’ test

---

<sup>4</sup>We also estimate specifications that include the teacher’s “raw” assessment while simultaneously controlling for students’ actual test scores.

performance. Consistent with previous findings, we observe the largest misalignment at level one, while differences are smaller at level three.

In summary, the data indicate that teacher predictions often deviate from test results, suggesting that teachers may possess significant private information about students’ skills. Next, we provide evidence to support this interpretation.

### 3 Do Teachers Have Private Information?

This section examines whether teachers hold private information that improves predictions of student performance. To explore this, we estimate empirical models aligned with the first stage of conventional value-added models (Chetty et al., 2014a; Koedel and Rockoff, 2015; Rothstein, 2017), while additionally incorporating our proxy for teacher private information (TPI) into the specifications. Specifically, our empirical model is given by:

$$\begin{aligned}
 Score_{ijt}^k = & \beta_0 + \beta_1 Score_{i,j',t-1}^k + \beta_2 (Score_{i,j',t-1}^k)^2 + \beta_3 (Score_{i,j',t-1}^k)^3 \\
 & \dots + \beta_4 TPI_{i,j',t-1}^k + \beta_5 (TPI_{i,j',t-1}^k)^2 + \mathbf{X}_{i,t}\boldsymbol{\gamma} + \mu_j + \varepsilon_{ijt}^k
 \end{aligned} \tag{1}$$

where  $Score_{ijt}^k$  represents the standardized test score of student  $i$  in period  $t$  in subject  $k$  with teacher  $j$ , standardized by year and grade. The term  $Score_{i,j',t-1}^k$  corresponds to the prior-year test score of student  $i$ , accounting for baseline academic performance. The variable  $TPI_{i,j',t-1}^k$  captures our measure of teacher private information (most of our analysis is based on the continuous measure, though we also present results using the discrete version and the raw teacher assessment variable), as reported by teacher  $j'$  in period  $t - 1$  for student  $i$ . In most cases, the reporting teacher  $j'$  differs from the current teacher  $j$  (i.e.,  $j' \neq j$ ). For example, our specification investigates whether the private information held by a student’s 4th-grade teacher ( $j'$ ) explains their performance in 5th grade. The vector  $\mathbf{X}_{i,t}$  includes student-level control variables in addition to the aggregates of these variables

at the classroom, school-year, and school-grade levels following Rothstein (2017). The term  $\mu_j$  represents the fixed effect of the current teacher  $j$ , capturing any time-invariant teacher-specific factors that may influence student performance. Finally,  $\varepsilon_{ijt}$  is the error term, which accounts for unobserved factors affecting student test scores. The empirical model is estimated separately for each subject  $k \in \{\text{math, reading}\}$ .

This specification allows us to assess whether our measure of teachers’ private information about students’ skills, as reported in a prior year, contributes to explaining subsequent academic performance. If the coefficients on teacher private information,  $\beta_4$  and  $\beta_5$ , are statistically significant, economically relevant, and contribute to an increase in the adjusted  $R^2$  of our empirical model, we can infer that our coarse measure of teacher private information plays a meaningful role in assessing students’ academic skills, strongly suggesting that teachers possess insights about student performance that are not fully captured by standardized test scores or other observable characteristics.

Table 2 reports estimation results for student performance in math and reading. Columns (1)–(2) present estimates from the first stage of conventional value-added models for math, controlling for prior-year test scores and teacher, year, and grade fixed effects. The two specifications differ in their included controls: Column (2) extends the specification in Column (1) by adding student demographics (gender and race), socioeconomic indicators (economic disadvantage), indicators of whether the course or test is repeated, skill-level indicators (special education and English learner status), and classroom-, school-year-, and school-grade-level aggregates of these variables, in addition to lagged scores, following Rothstein (2017). Columns (5)–(6) provide analogous estimates for reading performance. We find that incorporating these additional covariates increases the adjusted  $R^2$  of the models by approximately 3.6% and 5.7% for math and reading, respectively.

Columns (3)–(4) extend the specifications in Columns (1)–(2) for math, and Columns (7)–(8) extend those for reading, by incorporating our continuous measure of teacher private information (TPI), as defined in Section 2. Specifically, TPI is defined as the difference

(and squared difference) between the median test score corresponding to the teacher’s stated proficiency level at  $t - 1$  and the student’s actual EOG test performance at  $t - 1$ .<sup>5</sup> We find that including TPI significantly improves the predictive power of the models. First, the coefficients on TPI are significant and large in magnitude. Second, the estimated coefficients on prior test score level increase by about 13% for math and 17% for reading, suggesting the presence of a negative correlation between TPI and lagged scores.<sup>6</sup> Third, incorporating TPI enhances the adjusted  $R^2$  relative to conventional models, underscoring its explanatory value for test score variation. Specifically, Columns (3) and (7), which include TPI alongside only lagged test scores, yield adjusted  $R^2$  values only slightly lower than specifications controlling for gender, race, economically disadvantaged status, special education, English learner status, and variables capturing classroom and school characteristics (see Columns 2 and 6). This finding suggests that teacher private information is as influential as many student characteristics in predicting performance. To illustrate this point more rigorously, we decompose the  $R^2$  into the contributions of groups of regressors using Shapley values. Figure 1 shows, for example, that the contribution of lagged teacher private information ( $\text{TPI}_{t-1}$ ) to the  $R^2$  in math ranks third among all covariates (only behind the lagged scores), being its explanatory power larger than the combined explanatory power of gender, race, special education and English learner statuses, and economic disadvantage status. In reading, the contribution of TPI remains similarly substantial, nearly matching that of the combined student-level background characteristics, and is, if anything, even slightly larger than in math.

In summary, the analysis has shown so far that TPI, which is a coarse measure of teacher private information, significantly predicts student test scores, even when controlling for a comprehensive set of covariates. In particular, it meaningfully explains test score variation,

---

<sup>5</sup>Appendix Tables A5 and A6 present analogous estimates using alternative definitions of TPI: the discrete version of TPI and indicators based solely on the teacher’s prediction (rather than the difference between prediction and score level), respectively. The results are very similar regardless of how TPI is constructed.

<sup>6</sup>In fact, we do find that lagged scores and TPI are negatively correlated, with a correlation of -0.530 (-0.578) in math (reading). Appendix Table A7 shows that the effects of TPI persist even after including indicator variables for students that are close to the cut-off level values. This suggests that students right above or below the threshold level values are not driving the results.

as reflected in the increase in adjusted  $R^2$  compared to conventional value-added models, helps to provide a more accurate estimate of the effect of prior test scores on current performance, and its explanatory power rivals that of traditional student-level characteristics, underscoring its potential value in refining predictions of student performance. These findings strongly suggest the importance of incorporating teachers' private information into value-added models to better account for factors influencing student outcomes.

## 4 For Which Students Is Teacher Private Information More Relevant?

Previous research indicates that standardized test scores often underestimate the abilities of economically disadvantaged students and students of color (Rothstein, 2004; Geiser and Santelices, 2007). Motivated by these findings, we examine which student characteristics are most associated with discrepancies between teacher assessments and standardized test scores.

To investigate this, we regress an indicator for whether a teacher's level assessment at time  $t - 1$  deviates from a student's actual standardized test performance level on student characteristics. Table 3 reports OLS regression results, with Columns (1) and (2) corresponding to math and reading, respectively. Consistent with existing literature, our analysis reveals that teacher assessments diverge more markedly from standardized scores for certain student groups. In particular, we find significant positive coefficients for Black and economically disadvantaged students in both math and reading.<sup>7</sup> These discrepancies likely reflect teacher awareness of student abilities not captured by standardized tests due to external factors such as economic hardship, family instability, or other socio-economic challenges.

---

<sup>7</sup>It is important to mention that 65.4% of the total discrepancies concerning Black students and 54.2% of those concerning economically disadvantaged students in math correspond to teachers predicting a better performance than what was reflected in the test. While approximately 50% of the total discrepancies concerning economically disadvantaged and Black students in reading correspond to teachers predicting a better performance than what was reflected in the test.

Additionally, we also find significant discrepancies for students classified as special education or limited English proficient. This suggests that language barriers and specialized learning needs may prevent standardized tests from reflecting the actual skills of these students. Furthermore, students with lower performance on prior-year EOG exams ( $t - 2$ ) tend to experience greater discrepancies in assessments. Specifically, students scoring half a standard deviation above the mean in math (reading) have approximately a 4.50 (5.35) percentage points lower likelihood of discrepancy compared to those scoring at the mean level. This difference represents about 11 (12)% of the mean discrepancy rate.

Overall, the findings presented in the preceding and current sections indicate that teachers' private information plays a relevant role in assessing students whose standardized test scores tend to misrepresent their actual skill levels. These results also strongly suggest the importance of including teacher assessments into value-added models to better capture student performance and ensure fairer evaluations of teachers.

## 5 Does Teacher Private Information Impact Value-Added Estimates?

To illustrate the importance of including teacher private information (TPI) in value-added (VA) models, we performed three different analyses. First, following the two-step procedure established by [Chetty et al. \(2014a\)](#), we computed VA separately for math and reading, comparing the estimates obtained with and without TPI as a regressor in the first-stage estimation.<sup>8</sup> Second, we regress the teacher value-added estimates ( $VA_{j,t}$ ), obtained *without* incorporating TPI, on the classroom average  $TPI_{j',t-1}$  for teacher  $j$  at time  $t$ . This exercise helps assess whether selection on observables may contribute to differences in VA estimates.<sup>9</sup> Then, we estimate forecast bias following the approach in [Chetty et al. \(2014a\)](#)-section IV.

---

<sup>8</sup>Our estimation follows the first-stage specification detailed in Section 3.

<sup>9</sup>Appendix B outlines how omitting TPI from the first stage can bias VA when skills are measured with error.

Finally, we analyze how the inclusion of TPI affects teacher rankings based on VA estimates. Evaluating the sensitivity of teacher rankings to the inclusion of TPI is critical from a policy perspective. Influential research by [Chetty et al. \(2014a\)](#) and [Hanushek \(2009\)](#) suggests replacing the lowest-performing 5% of teachers to enhance student outcomes. However, accurately identifying this group becomes challenging if VA estimates vary depending on the model specification.

Appendix Figure [A1](#) quantifies the absolute value difference in individual teachers' VA estimates resulting from the inclusion and exclusion of TPI, expressing each absolute change relative to the overall yearly standard deviation of VA. In math, 71% of teachers experience a shift exceeding 5% of the standard deviation, while 46% experience changes greater than 10%. For reading, these figures are 79% and 60%, respectively. While the imprecision of value-added estimates could drive some of these differences, Table [A8](#) shows that average  $TPI_{j',t-1}$  at the classroom level negatively correlates with teacher VA when TPI is not included when recovering VA estimates. More specifically, the coefficient on average  $TPI_{j',t-1}$  is -0.035 (-0.012) and statistically significant at the 1% level when the dependent variable is math (reading) VA. Finally, we followed [Chetty et al. \(2014a\)](#) to estimate forecast bias in value-added due to the omission of TPI in estimation.<sup>10</sup> Table [A9](#) shows that the magnitude of forecast bias is approximately 1.1% and statistically significant for math, and 3.1% and statistically significant for reading. These forecast biases are more than five times and ten times larger for math and reading, respectively, than the one found by [Chetty et al. \(2014a\)](#) where they consider the following observables not included in standard VA models: mother's age at child's birth, indicators for parent's 401(k) contributions and home ownership, and an indicator for the parent marital status interacted with a quartic in parent household income. Given that our coarse measure of teacher private information can impact value-

---

<sup>10</sup>To estimate the forecast bias, we implemented the following steps for each subject: 1) regress  $TPI_{i,j',t-1}$  on the same set of controls as in eq.(1) except for TPI and recover the residuals ( $r1$ ), 2) Then regress the residuals from the first-stage of VA estimation (without TPI) on  $r1$ , 3) calculate predicted values from that regression ( $\rho * r1$ ), 4) regress  $\rho * r1$  on value-added estimates that do not include TPI in the first stage of their estimation.

added estimates, our findings indicate the likely importance of incorporating more refined measures of teacher private information when assessing teacher effectiveness.

Finally, to illustrate the practical implications of including TPI in value-added models, Figure 2 shows how teachers move across VA decile ranks when value-added estimates are computed with and without TPI in the first step of the estimation. We performed this analysis only for VA estimates corresponding to the year 2009 to avoid repeating the same teacher across years. After integrating TPI into the model, approximately 10% of the teachers initially ranked in the lowest decile for math no longer remain in that group, while for reading, it is at 14%. Among those in the bottom 5%, only 67.4% in math and 63% in reading *remained* in the bottom 5% after including TPI.<sup>11</sup> Similar patterns are observed at the top of the VA distribution, an important finding given that teacher bonuses are sometimes awarded to top-performing teachers based on value-added estimates (Dee and Wyckoff, 2015). In summary, these findings imply that the composition of teachers identified for potential dismissal or targeted salary bonuses can be influenced by the inclusion of teacher private information.

To conclude, our results underscore that including TPI in VA models provides a more robust foundation for policy decisions targeting the lowest- and highest-performing teachers. By reducing the risk of misclassification, TPI can help administrators make fairer and more impactful choices, ultimately supporting better outcomes for students. Additionally, the fact that our measure of TPI is coarse suggests that incorporating more detailed teacher information about students has the potential to alter value-added estimates even more.

---

<sup>11</sup>Given that imprecise estimates could drive changes in teacher rankings, we estimated value-added models where TPI was randomly allocated across students within the classroom to preserve the same mean and standard deviation of TPI within class. Figure A2 shows that the VA decile rankings computed without TPI are nearly identical to those computed including randomized TPI (within class), with more than 99% overlap in each decile for both math and reading. This similarity underscores that the actual TPI of each student contains meaningful information beyond random noise.

## 6 Conclusions

This paper examines the implications of integrating teachers' private information into value-added models for educational assessment. Our analysis reveals that TPI enhances the predictive accuracy of student achievement beyond traditional control variables such as demographics and previous test scores. Notably, discrepancies between teacher and test levels are more prevalent among Black students, economically disadvantaged students, English learners, and students in special education, groups typically underserved by standardized assessments. Additionally, we show that TPI correlates with VA estimates when not included in estimation, suggesting a problem of selection on observables. Finally, neglecting TPI can lead to misclassification in teacher effectiveness rankings. Thus, accounting for TPI not only sharpens the identification of consistently underperforming teachers but also promotes equity by more accurately reflecting the skill levels of the student populations. Our findings indicate that incorporating and developing more refined measures of teacher private information into VA models could improve their effectiveness in evaluating teacher performance.

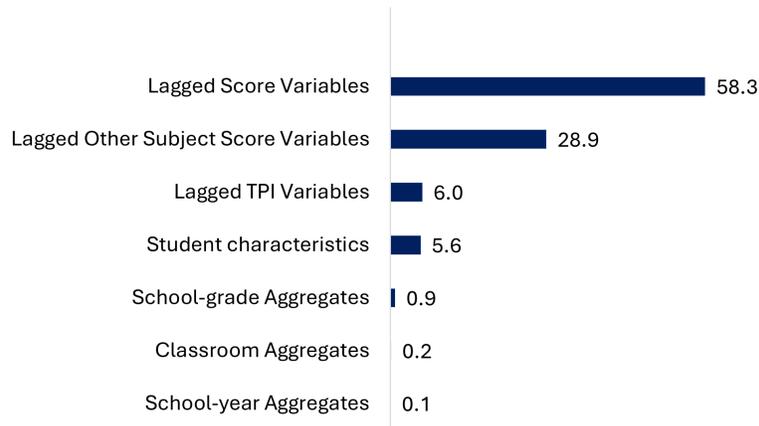
## References

- Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James Wyckoff**, “Measuring test measurement error: The importance of test reliability for estimates of teacher effectiveness,” *Educational Researcher*, 2013, *42* (6), 308–317.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff**, “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates,” *American Economic Review*, 2014, *104* (9), 2593–2632.
- , – , **and** – , “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood,” *American Economic Review*, 2014, *104* (9), 2633–2679.
- Dee, Thomas S and James Wyckoff**, “Incentives, selection, and teacher performance: Evidence from IMPACT,” *Journal of Policy Analysis and Management*, 2015, *34* (2), 267–297.
- Geiser, Saul and Maria Veronica Santelices**, “Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes,” 2007.
- Hanushek, Eric A.**, “Teacher deselection,” in “Creating a New Teaching Profession,” National Center on Performance Incentives, 2009, pp. 165–180.
- Kane, Thomas J. and Douglas O. Staiger**, “The promise and pitfalls of using imprecise school accountability measures,” *Journal of Economic Perspectives*, 2002, *16* (4), 91–114.
- Koedel, Cory and Jonah E Rockoff**, “Value-added modeling: A review,” *Economics of Education Review*, 2015, *47*, 180–195.
- Rothstein, Jesse**, “Measuring the impacts of teachers: Comment,” *American Economic Review*, 2017, *107* (6), 1656–1684.

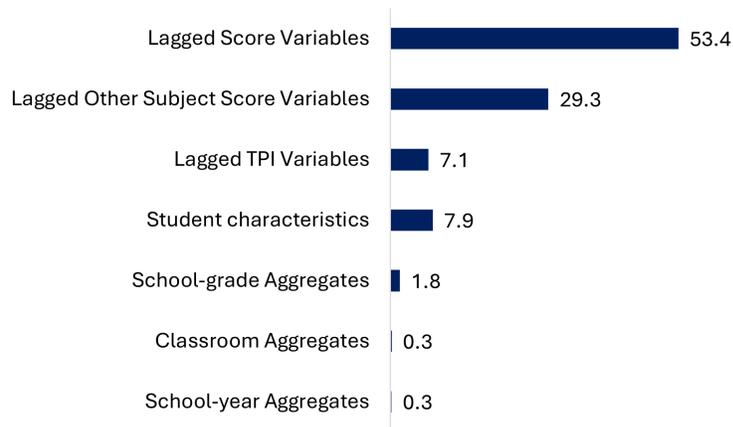
**Rothstein, Jesse M**, “College performance predictions and the SAT,” *Journal of Econometrics*, 2004, 121 (1-2), 297–317.

## 7 Figures

Figure 1: Group Contribution to  $R^2$



(a) Math



(b) Reading

Note: This figure displays the share of explained variance (R-squared) in current student test scores in Math (Panel (a)) and Reading (Panel (b)) EOG exams attributed to different groups of explanatory variables, based on Shapley value decompositions. Lagged Score Variables include the previous year's test scores in the respective subject, in addition to the squared and cubed terms. Lagged Other Subject Score Variables include these variables for the other subject, i.e., reading for Panel (a) and math for Panel (b). Teacher Private Information (TPI) is defined as the difference between a student's standardized score and the median standardized score of the achievement level matched to the teacher's predicted achievement level for the student, and Lagged TPI Variables include both the lagged TPI and its squared term. Student characteristics are the gender, ethnicity, Special Education, economic disadvantage, and English learner statuses (ELS) of the student. School-grade, Classroom, and School-year aggregates include the aggregates for all of the remaining variables at the indicated level. Each bar reflects the fraction of the total explained variance attributable to that group, conditional on the other variables included in the model. Overall  $R^2$  is 0.726 for Math and 0.682 for Reading.

Figure 2: Percentage of VA Estimate Deciles with  $TPI_{t-1}$  by Deciles without  $TPI_{t-1}$

		Decile of VA Estimate with TPI									
		1	2	3	4	5	6	7	8	9	10
Decile of VA Estimate without TPI	1	89.9	9.9	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	9.9	74.4	15.1	0.6	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.2	15.1	64.6	19.7	0.5	0.0	0.0	0.0	0.0	0.0
	4	0.0	0.6	19.6	57.9	21.2	0.6	0.0	0.0	0.0	0.0
	5	0.0	0.0	0.5	21.2	57.2	19.5	1.6	0.0	0.0	0.0
	6	0.0	0.0	0.0	0.6	20.6	59.6	17.9	1.3	0.0	0.0
	7	0.0	0.0	0.0	0.0	0.5	19.6	61.4	17.7	0.8	0.0
	8	0.0	0.0	0.0	0.0	0.0	0.6	19.2	67.6	12.6	0.0
	9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.4	78.1	8.5
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.5	91.5

(a) Math

		Decile of VA Estimate with TPI									
		1	2	3	4	5	6	7	8	9	10
Decile of VA Estimate without TPI	1	86.3	13.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	13.4	63.8	18.7	3.6	0.5	0.0	0.0	0.0	0.0	0.0
	3	0.3	20.6	54.6	21.5	2.2	0.8	0.0	0.0	0.0	0.0
	4	0.0	1.9	22.8	46.9	23.4	4.6	0.5	0.0	0.0	0.0
	5	0.0	0.2	3.8	24.4	46.5	21.1	3.6	0.5	0.0	0.0
	6	0.0	0.0	0.0	3.3	22.3	47.8	23.1	3.1	0.3	0.0
	7	0.0	0.0	0.0	0.3	4.7	22.6	45.4	24.0	2.8	0.2
	8	0.0	0.0	0.0	0.0	0.3	3.1	26.4	50.2	19.8	0.2
	9	0.0	0.0	0.0	0.0	0.0	0.0	1.1	21.9	62.1	14.9
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	14.9	84.8

(b) Reading

Note: Each square indicates the fraction of teachers who remain in the same decile of their estimated value-added (VA) ranking when moving from a specification that excludes the lagged Teacher Private Information ( $TPI_{t-1}$ , rows) to one that includes it (columns) for the given subject, Math and Reading in Panels (a) and (b), respectively for the year 2009. VA measures a teacher's contribution to student test-score growth, estimated via controlling for prior test scores, student and classroom covariates.  $TPI_{t-1}$  is defined as the difference between a student's lagged standardized score and the median standardized score of the achievement level matched to the previous year's teacher's judgment. Deciles are determined by sorting teachers from lowest to highest VA in each specification.

# Tables

Table 1: Summary Statistics of TPI

	(1)	(2)	(3)	(4)	(5)	(6)
	N	Share (%)	Continuous TPI Variable			
			Mean	SD	Min	Max
<i>Panel A: Math TPI Levels</i>						
-3	164	0.0	-3.316	0.468	-4.945	-2.722
-2	6,834	0.9	-1.941	0.403	-3.909	0.155
-1	137,229	18.0	-0.984	0.373	-2.671	1.190
0	462,054	60.7	-0.049	0.419	-1.660	2.497
1	145,521	19.1	0.993	0.325	-0.353	2.689
2	9,384	1.2	2.064	0.290	0.920	3.572
3	184	0.0	3.143	0.242	2.822	4.087
<i>Panel B: Reading TPI Levels</i>						
-3	493	0.1	-2.940	0.404	-4.771	-2.343
-2	13,324	1.8	-1.888	0.365	-4.209	0.222
-1	176,863	23.3	-0.948	0.388	-3.087	2.000
0	433,359	57.0	-0.072	0.431	-1.988	2.197
1	122,843	16.2	0.939	0.318	-1.063	3.007
2	13,288	1.8	1.816	0.371	-0.528	3.828
3	388	0.1	2.714	0.417	1.889	4.323

Note: Table summarizes two versions of our lagged Teacher Private I Information (TPI) measure. The first is a discrete measure from -3 to +3, defined as the teacher's predicted achievement level minus the student's actual level. The second is a continuous measure that quantifies the difference between the student's actual standardized EOG score and the median standardized score corresponding to the teacher's predicted level. Panel A focuses on math TPI, while Panel B presents reading TPI. Each row corresponds to a given discrete mismatch (-3 to +3), and columns (1) through (6) display the number of observations (N), the share of the total sample in percentage, and the mean, standard deviation, minimum, and maximum values of the continuous TPI.

Table 2: Relevance of  $TPI_{t-1}$  Across Alternate First-Stage Value-Added Specifications

	Math				Reading			
	Without $TPI_{t-1}$		With $TPI_{t-1}$		Without $TPI_{t-1}$		With $TPI_{t-1}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score $_{t-1}$	0.856*** (0.001)	0.711*** (0.002)	0.928*** (0.001)	0.802*** (0.002)	0.840*** (0.001)	0.665*** (0.002)	0.942*** (0.001)	0.778*** (0.002)
Score $^2_{t-1}$	-0.011*** (0.001)	-0.011*** (0.001)	-0.005*** (0.001)	-0.009*** (0.001)	-0.028*** (0.001)	-0.031*** (0.000)	-0.017*** (0.001)	-0.026*** (0.001)
Score $^3_{t-1}$	-0.021*** (0.000)	-0.019*** (0.000)	-0.016*** (0.000)	-0.016*** (0.000)	-0.026*** (0.000)	-0.024*** (0.000)	-0.019*** (0.000)	-0.021*** (0.000)
$TPI_{t-1}$			0.213*** (0.001)	0.172*** (0.001)			0.251*** (0.001)	0.183*** (0.001)
$TPI^2_{t-1}$			0.006*** (0.001)	0.005*** (0.001)			0.000 (0.001)	0.002*** (0.001)
Female		-0.030*** (0.001)		-0.030*** (0.001)		0.047*** (0.001)		0.033*** (0.001)
Black		-0.105*** (0.003)		-0.101*** (0.003)		-0.0650*** (0.003)		-0.0627*** (0.003)
Hispanic		-0.005 (0.003)		-0.003 (0.003)		-0.004 (0.004)		0.000 (0.004)
Other race (non-White)		-0.023*** (0.003)		-0.020*** (0.002)		0.026*** (0.003)		0.026*** (0.003)
Economically Disadvantaged		-0.035*** (0.001)		-0.027*** (0.001)		-0.038*** (0.001)		-0.032*** (0.001)
Special Education		-0.070*** (0.004)		-0.043*** (0.004)		-0.126*** (0.005)		-0.078*** (0.004)
English Learner		0.024*** (0.004)		0.033*** (0.003)		-0.093*** (0.004)		-0.070*** (0.004)
Constant	0.016*** (0.003)	1.891*** (0.146)	-0.041*** (0.003)	1.605*** (0.143)	0.027*** (0.003)	1.149*** (0.159)	0.107*** (0.003)	0.971*** (0.156)
Additional controls		✓		✓		✓		✓
Classroom & School Char.s		✓		✓		✓		✓
Observations	761,370	761,370	761,370	761,370	760,558	760,558	760,558	760,558
Adjusted R-squared	0.635	0.658	0.657	0.672	0.577	0.610	0.603	0.622

Note: Table presents the coefficients from a regression where the outcome variables are the standardized math and reading test scores. Lagged Teacher Private Information ( $TPI_{t-1}$ ) is the difference between a student's lagged standardized score and the median score of the achievement level corresponding to the previous year's teacher's judgment on the student's achievement in Math (columns (1)-(4)) and Reading (columns (5)-(8)). The baseline ethnicity is white. Other ethnicities are Asian, Native American, and multiethnic. Additional controls include any pre-existing scores for the students across both subjects, in addition to dummy variables for repeated grade and repeated test. Classroom and school characteristics include all of the characteristics for the students pooled at the classroom, school-year, and school-grade levels. All specifications account for teacher, year, and grade fixed effects. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

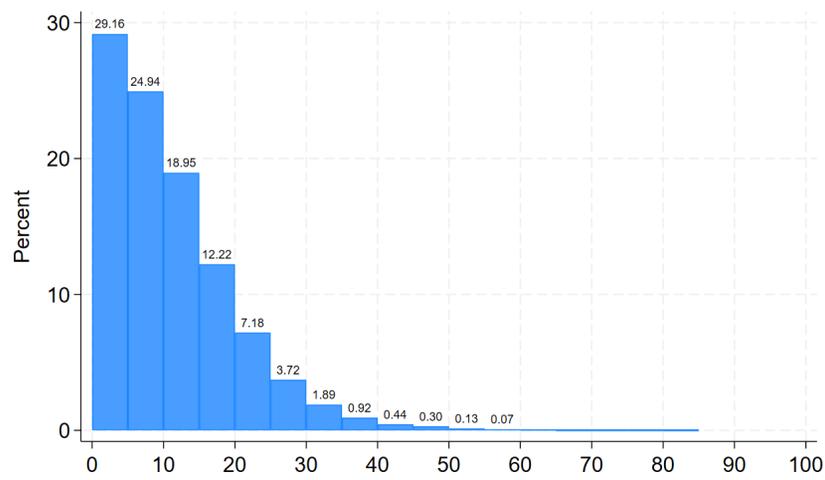
Table 3: Characteristics' Effects on Binary  $TPI_{t-1}$ 

	Math (1)	Reading (2)
Score $_{t-2}$	-0.079*** (0.001)	-0.093*** (0.001)
Score $^2_{t-2}$	-0.022*** (0.001)	-0.028*** (0.001)
Female	-0.006*** (0.002)	-0.012*** (0.002)
Black	0.016*** (0.003)	0.013*** (0.003)
Hispanic	-0.003 (0.004)	0.007* (0.004)
Other race (non-White)	-0.013*** (0.003)	-0.013*** (0.004)
Economically Disadvantaged	0.012*** (0.002)	0.018*** (0.002)
Special Education	0.041*** (0.006)	0.047*** (0.006)
English Learner	0.015*** (0.005)	0.014*** (0.006)
Constant	0.413*** (0.002)	0.446*** (0.002)
Mean of Dependent Var.	0.394	0.431
Observations	280,078	279,806
Adjusted R-squared	0.070	0.070

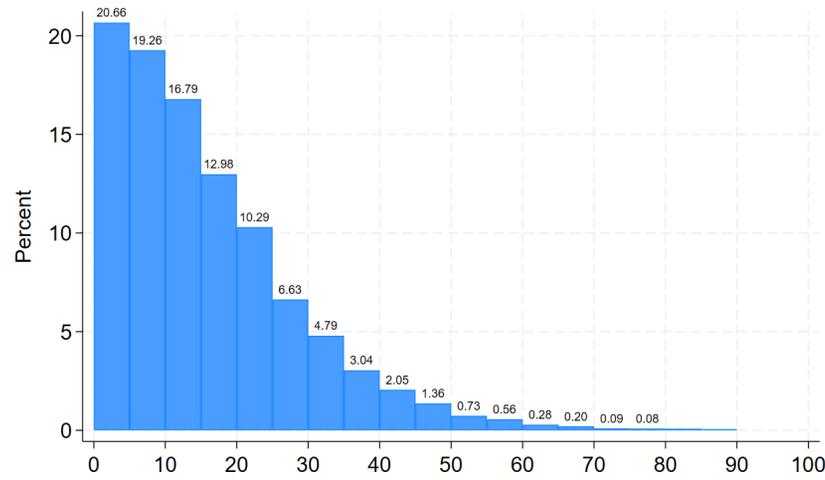
Note: This table displays the coefficients from a regression where the dependent variable is an indicator that takes the value one if the lagged Teacher Private Information ( $TPI_{t-1}$ ) prediction in levels differs from the achievement level of the student in the previous year in the given subject. Math and Reading in columns 1 and 2, respectively. The baseline ethnicity is white. Other ethnicities are Asian, Native American, and multiethnic. Other controls not included in the table are dummy variables indicating whether the student is repeating the grade and/or the test. Both columns account for lagged teacher fixed effects. The number of observations is lower relative to Table 2 due to missing twice-lagged scores for all of the fourth-grade students, and a small proportion of the fifth-grade students (6.1%) who were not tested in their third grade, potentially due to not residing in North Carolina. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

# A Online Appendix

Figure A1: Distribution of Absolute Changes in Value-Added Estimates After Including TPI as % of Overall VA Std. Dev.



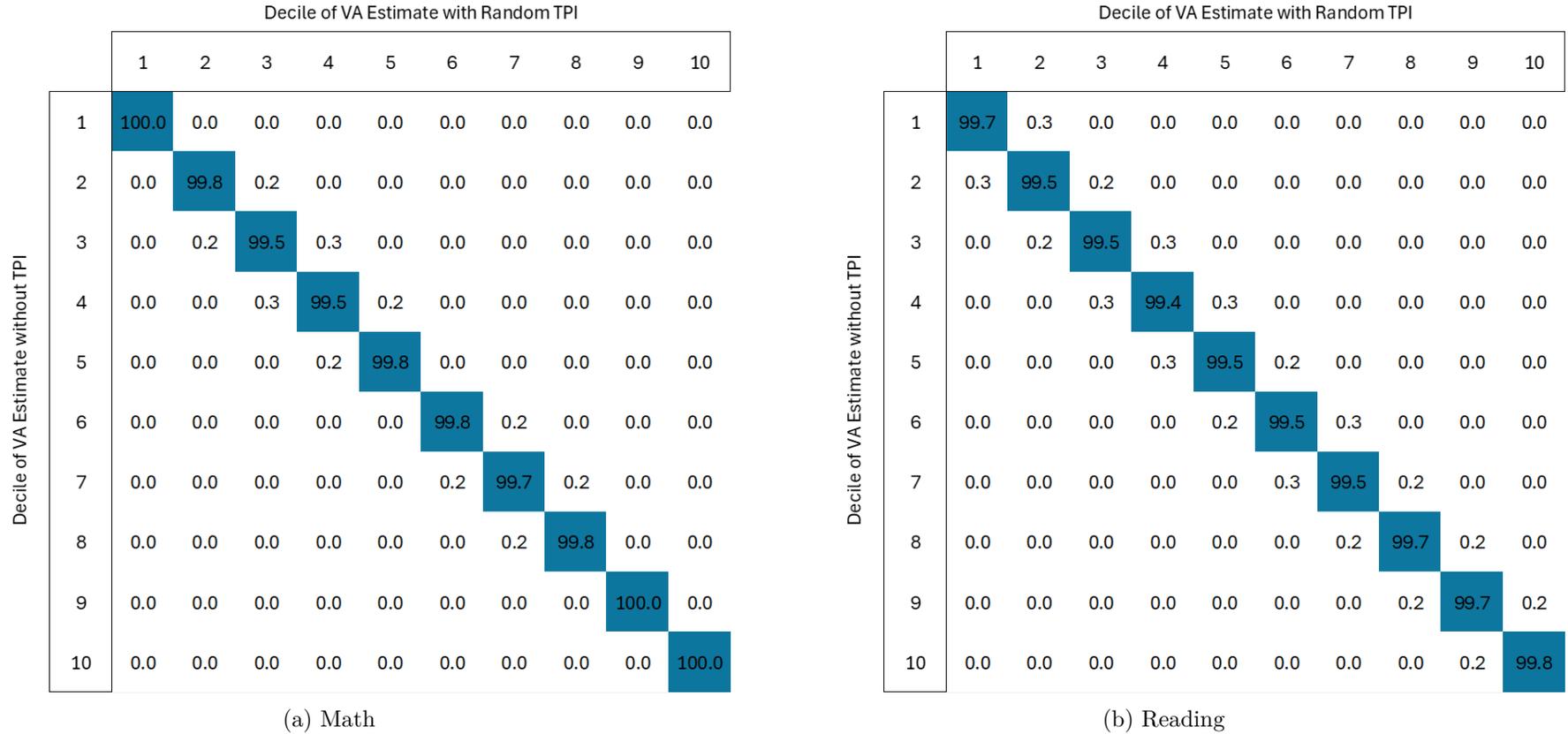
(a) Math



(b) Reading

Note: The figure presents by subject histograms of the distribution of the absolute difference between the value-added estimates with and without TPI, scaled by the subject-year standard deviation of the value-added estimates without TPI. The numbers at the top of the bars indicate the proportion of VA estimates that fall in that group.

Figure A2: Percentage Overlap of VA Estimate Deciles when VA is estimated without  $TPI_{t-1}$  and with Randomized  $TPI_{t-1}$



Note: Each square indicates the fraction of teachers who remain in the same decile of their estimated value-added (VA) ranking when moving from a specification that includes a randomized lagged Teacher Private Information ( $TPI_{t-1}$ , columns) within the classroom on the same subject to one that does not include lagged Teacher Private Information (rows) for the given subject, Math and Reading in Panels (a) and (b), respectively for the year 2009. VA measures a teacher's contribution to student test-score growth, estimated via controlling for prior test scores, student and classroom covariates.  $TPI_{t-1}$  is defined as the difference between a student's lagged standardized score and the median standardized score of the achievement level matched to the previous year's teacher's judgment. Deciles are determined by sorting teachers from lowest to highest VA in each specification.

Table A1: Summary Statistics

VARIABLES	Mean	Std. Dev.
<i>Panel A: Student-Year (N=762,108)</i>		
Female	0.499	0.500
White	0.585	0.493
Black	0.212	0.408
Hispanic	0.093	0.291
Asian	0.023	0.150
Native	0.013	0.112
Multiethnic	0.074	0.262
ELS	0.052	0.223
EDS	0.294	0.455
Special Education	0.025	0.155
<i>Panel B: Teacher-Year (N=37,336)</i>		
Female	0.899	0.302
White	0.871	0.335
Black	0.110	0.312
Hispanic	0.006	0.074
Asian	0.004	0.066
Native	0.008	0.087
Multiethnic	0.002	0.046

Note: Summary statistics include observations corresponding to 4th and 5th grades for the years 2007 through 2011. Panel A shows student-year observations, and Panel B shows teacher-year observations. Each teacher-year is matched to a classroom. Observations are limited to the ones having lagged achievement levels and teacher judgments for the purposes of our analyses. ELS: English learner status. EDS: economically disadvantaged status.

Table A2: Summary Statistics of TPI: Excluding observations that are close to the EOG Exam level thresholds

	(1)	(2)	(3)	(4)	(5)	(6)
	N	Share (%)	Continuous TPI Variable			
			Mean	SD	Min	Max
<i>Panel A: Math TPI Levels</i>						
-3	114	0.0	-3.429	0.513	-4.945	-2.722
-2	4,165	0.5	-1.989	0.448	-3.909	-0.305
-1	87,884	11.5	-1.059	0.399	-2.671	0.690
0	306,485	40.3	-0.131	0.450	-1.660	2.497
1	95,215	12.5	0.903	0.321	-0.353	2.689
2	7,588	1.0	1.988	0.257	0.920	3.572
3	168	0.0	3.099	0.203	2.822	4.087
<i>Panel B: Reading TPI Levels</i>						
-3	410	0.1	-3.016	0.401	-4.771	-2.445
-2	8,796	1.2	-1.995	0.375	-4.209	-1.041
-1	132,304	17.4	-1.019	0.400	-3.087	2.000
0	331,383	43.6	-0.152	0.441	-1.988	2.197
1	85,210	11.2	0.891	0.337	-1.063	2.862
2	9,115	1.2	1.841	0.411	-0.528	3.629
3	253	0.0	2.772	0.469	1.889	4.323

Note: Table summarizes two versions of our lagged Teacher Private Information (TPI) measure, excluding students with standardized scores that are within half a standard deviation of the threshold of the achievement levels. The first is a discrete measure from -3 to +3, defined as the teacher's predicted achievement level minus the student's actual level. The second is a continuous measure that quantifies the difference between the student's actual standardized EOG score and the median standardized score corresponding to the teacher's predicted level. Panel A focuses on math TPI, while Panel B presents reading TPI. Each row corresponds to a given discrete mismatch (-3 to +3), and columns (1) through (6) display the number of observations (N), the share relative to the total number of observations in the sample, not just those far from the thresholds, and the mean, standard deviation, minimum, and maximum values of the continuous TPI.

Table A3: Teacher Judgments by Student Achievement Levels

	(1)	(2)	(3)	(4)	(5)
	N	Mean	SD	Min	Max
<i>Math achievement levels:</i>					
1	29,589	1.942	0.716	1	4
2	114,789	2.424	0.694	1	4
3	378,421	3.048	0.618	1	4
4	238,571	3.659	0.503	1	4

*Reading achievement levels:*

1	40,414	2.028	0.732	1	4
2	85,648	2.460	0.704	1	4
3	327,101	2.998	0.642	1	4
4	307,395	3.562	0.557	1	4

Note: This table shows the average teacher judgment levels (defined on a 1 to 4 discrete scale) conditional on the students' performance level in the EOG exams. Panel A focuses on math, while Panel B presents average reading judgment levels. Each row corresponds to a given achievement level for the students, and columns (1) through (5) display the number of observations (N), and the mean, standard deviation, minimum, and maximum values of the teacher judgment level.

Table A4: Continuous TPI Measure Statistics by Student Achievement Levels

	(1)	(2)	(3)	(4)	(5)
	N	Mean	SD	Min	Max
<i>Math achievement levels:</i>					
1	29,589	0.961	0.773	-1.191	4.087
2	114,789	0.522	0.743	-2.101	3.153
3	378,421	0.060	0.674	-3.125	2.521
4	238,571	-0.496	0.624	-4.945	1.083
<i>Reading achievement levels:</i>					
1	40,414	1.038	0.722	-1.354	4.323
2	85,648	0.488	0.697	-1.862	3.828
3	327,101	0.021	0.654	-3.142	2.048
4	307,395	-0.570	0.636	-4.771	1.129

Note: This table shows the average continuous TPI measure conditional on students' actual test score level performance, reflecting the gap between the median standardized score corresponding to the teacher's predicted level and the student's actual standardized EOG score. Panel A focuses on math, while Panel B presents reading judgment levels. Each row corresponds to a given achievement level for the students, and columns (1) through (5) display the number of observations (N), and the mean, standard deviation, minimum, and maximum values of the teacher judgment level.

Table A5: Relevance of the Discrete TPI Variable Across Alternate First-Stage Value-Added Specifications

	Math				Reading			
	Without TPI		With TPI		Without TPI		With TPI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score <sub>t-1</sub>	0.856*** (0.001)	0.711*** (0.002)	0.887*** (0.001)	0.759*** (0.002)	0.840*** (0.001)	0.665*** (0.002)	0.903*** (0.001)	0.747*** (0.002)
Score <sub>t-1</sub> <sup>2</sup>	-0.011*** (0.001)	-0.011*** (0.001)	-0.0201*** (0.001)	-0.020*** (0.001)	-0.028*** (0.001)	-0.031*** (0.001)	-0.047*** (0.001)	-0.047*** (0.001)
Score <sub>t-1</sub> <sup>3</sup>	-0.021*** (0.000)	-0.019*** (0.000)	-0.021*** (0.000)	-0.020*** (0.000)	-0.026*** (0.000)	-0.024*** (0.000)	-0.033*** (0.000)	-0.031*** (0.000)
Discrete $TPI_{t-1} = -3$			-0.186*** (0.042)	-0.125*** (0.041)			-0.440*** (0.026)	-0.304*** (0.026)
Discrete $TPI_{t-1} = -2$			-0.331*** (0.007)	-0.254*** (0.006)			-0.403*** (0.005)	-0.273*** (0.005)
Discrete $TPI_{t-1} = -1$			-0.191*** (0.002)	-0.151*** (0.002)			-0.224*** (0.002)	-0.157*** (0.002)
Discrete $TPI_{t-1} = 1$			0.180*** (0.002)	0.144*** (0.002)			0.187*** (0.002)	0.136*** (0.002)
Discrete $TPI_{t-1} = 2$			0.400*** (0.006)	0.323*** (0.006)			0.364*** (0.006)	0.271*** (0.005)
Discrete $TPI_{t-1} = 3$			0.536*** (0.039)	0.421*** (0.038)			0.572*** (0.030)	0.436*** (0.029)
Female		-0.030*** (0.001)		-0.030*** (0.001)		0.047*** (0.001)		0.036*** (0.001)
Black		-0.105*** (0.003)		-0.102*** (0.003)		-0.0650*** (0.003)		-0.0630*** (0.003)
Hispanic		-0.005 (0.003)		-0.003 (0.003)		-0.004 (0.004)		-0.001 (0.004)
Other race (non-White)		-0.023*** (0.003)		-0.020*** (0.002)		0.026*** (0.003)		0.026*** (0.003)
Economically Disadvantaged		-0.035*** (0.001)		-0.029*** (0.001)		-0.038*** (0.001)		-0.034*** (0.001)
Special Education		-0.070*** (0.004)		-0.053*** (0.004)		-0.126*** (0.005)		-0.088*** (0.004)
English Learner		0.024*** (0.004)		0.031*** (0.004)		-0.093*** (0.004)		-0.075*** (0.004)
Constant	0.016*** (0.003)	1.891*** (0.146)	-0.012*** (0.003)	1.749*** (0.144)	0.027*** (0.003)	1.149*** (0.159)	0.104*** (0.003)	1.027*** (0.157)
Additional controls		✓		✓		✓		✓
Classroom & School Char.s		✓		✓		✓		✓
Observations	761,370	761,370	761,370	761,370	760,558	760,558	760,558	760,558
Adjusted R-squared	0.635	0.658	0.652	0.668	0.577	0.610	0.597	0.619

Note: Table presents the coefficients of the OLS regression where the outcome variables are the standardized math and reading test scores. Discrete  $TPI_{t-1}$  is the difference between the teacher's judgment of the student and the student's achievement level in Math (columns (1)-(4)) and Reading (columns (5)-(8)) on 4-point scales. The baseline ethnicity is white. Other ethnicities are Asian, Native American, and multiethnic. Additional controls include any pre-existing scores for the students across both subjects, in addition to dummy variables for repeated grade and repeated test. Classroom and school characteristics include all of the characteristics for the students pooled at the classroom, school-year, and school-grade levels. All specifications account for teacher, year, and grade fixed effects. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table A6: Relevance of  $TPI_{t-1}$  Across Alternate First-Stage Value-Added Specifications

	Math				Reading			
	Without $TPI_{t-1}$		With $TPI_{t-1}$		Without $TPI_{t-1}$		With $TPI_{t-1}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score $_{t-1}$	0.856*** (0.001)	0.711*** (0.002)	0.713*** (0.001)	0.628*** (0.002)	0.840*** (0.001)	0.665*** (0.002)	0.687*** (0.001)	0.594*** (0.002)
Score $^2_{t-1}$	-0.011*** (0.001)	-0.011*** (0.001)	-0.006*** (0.001)	-0.009*** (0.001)	-0.028*** (0.001)	-0.031*** (0.001)	-0.020*** (0.001)	-0.027*** (0.001)
Score $^3_{t-1}$	-0.021*** (0.000)	-0.019*** (0.000)	-0.015*** (0.000)	-0.016*** (0.000)	-0.026*** (0.000)	-0.024*** (0.000)	-0.019*** (0.000)	-0.021*** (0.000)
Prediction $_{t-1} = 2$			0.184*** (0.004)	0.145*** (0.004)			0.206*** (0.004)	0.151*** (0.004)
Prediction $_{t-1} = 3$			0.411*** (0.004)	0.325*** (0.004)			0.447*** (0.004)	0.328*** (0.004)
Prediction $_{t-1} = 4$			0.671*** (0.004)	0.537*** (0.004)			0.736*** (0.005)	0.536*** (0.005)
Female		-0.030*** (0.001)		-0.030*** (0.001)		0.047*** (0.001)		0.033*** (0.001)
Black		-0.105*** (0.003)		-0.101*** (0.003)		-0.065*** (0.003)		-0.063*** (0.003)
Hispanic		-0.005 (0.003)		-0.002 (0.003)		-0.004 (0.004)		0.000 (0.004)
Other race (non-White)		-0.023*** (0.003)		-0.020*** (0.002)		0.026*** (0.003)		0.026*** (0.003)
Economically Disadvantaged		-0.035*** (0.001)		-0.028*** (0.001)		-0.038*** (0.001)		-0.030*** (0.001)
Special Education		-0.070*** (0.004)		-0.041*** (0.004)		-0.126*** (0.005)		-0.089*** (0.004)
English Learner		0.024*** (0.004)		0.033*** (0.003)		-0.093*** (0.004)		-0.072*** (0.004)
Constant	0.016*** (0.003)	1.891*** (0.146)	-0.417*** (0.005)	1.308*** (0.143)	0.027*** (0.003)	1.149*** (0.159)	-0.444*** (0.005)	0.586*** (0.156)
Additional controls		✓		✓		✓		✓
Classroom & School Char.s		✓		✓		✓		✓
Observations	761,370	761,370	761,370	761,370	760,558	760,558	760,558	760,558
Adjusted R-squared	0.635	0.658	0.657	0.672	0.577	0.610	0.603	0.623

Note: Table presents the coefficients of a regression where the outcome variables are the standardized math and reading test scores. Lagged Teacher Prediction (Prediction $_{t-1}$ ) is the teacher's prediction on the student's achievement on a 4-point scale in Math (columns (1)-(4)) and Reading (columns (5)-(8)) in the previous year before the EOG exams. The baseline ethnicity is white. Other ethnicities are Asian, Native American, and multiethnic. Additional controls include any pre-existing scores for the students across both subjects, in addition to dummy variables for repeated grade and repeated test. Classroom and school characteristics include all of the characteristics for the students pooled at the classroom, school-year, and school-grade levels. All specifications account for teacher, year, and grade fixed effects. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table A7: Relevance of TPI Variable Across Alternate First-Stage Value-Added Specifications Including Indicators for Being Close to Threshold Levels

	Math				Reading			
	Without TPI		With TPI		Without TPI		With TPI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score <sub>t-1</sub>	0.863*** (0.001)	0.720*** (0.002)	0.928*** (0.001)	0.801*** (0.002)	0.845*** (0.001)	0.671*** (0.002)	0.942*** (0.001)	0.778*** (0.002)
Score <sub>t-1</sub> <sup>2</sup>	-0.011*** (0.001)	-0.011*** (0.001)	-0.006*** (0.001)	-0.010*** (0.001)	-0.027*** (0.001)	-0.031*** (0.001)	-0.017*** (0.001)	-0.026*** (0.001)
Score <sub>t-1</sub> <sup>3</sup>	-0.021*** (0.000)	-0.020*** (0.000)	-0.015*** (0.000)	-0.016*** (0.000)	-0.026*** (0.000)	-0.024*** (0.000)	-0.019*** (0.000)	-0.021*** (0.000)
TPI <sub>t-1</sub>			0.214*** (0.001)	0.173*** (0.001)			0.251*** (0.001)	0.183*** (0.001)
TPI <sub>t-1</sub> <sup>2</sup>			0.006*** (0.001)	0.006*** (0.001)			-0.000 (0.001)	0.002*** (0.001)
Female		-0.030*** (0.001)		-0.030*** (0.001)		0.046*** (0.001)		0.033*** (0.001)
Black		-0.105*** (0.003)		-0.101*** (0.003)		-0.065*** (0.003)		-0.063*** (0.003)
Hispanic		-0.005 (0.003)		-0.003 (0.003)		-0.004 (0.004)		0.000 (0.004)
Other race (non-White)		-0.022*** (0.003)		-0.019*** (0.002)		0.026*** (0.003)		0.026*** (0.003)
Economically Disadvantaged		-0.034*** (0.001)		-0.027*** (0.001)		-0.038*** (0.001)		-0.032*** (0.001)
Special Education		-0.068*** (0.004)		-0.044*** (0.004)		-0.123*** (0.005)		-0.078*** (0.004)
English Learner		0.025*** (0.004)		0.033*** (0.003)		-0.090*** (0.004)		-0.070*** (0.004)
Constant	0.002 (0.003)	1.867*** (0.146)	-0.039*** (0.003)	1.606*** (0.143)	0.029*** (0.003)	1.150*** (0.159)	0.107*** (0.003)	0.973*** (0.156)
Additional controls		✓		✓		✓		✓
Classroom & School Char.s		✓		✓		✓		✓
Observations	761,370	761,370	761,370	761,370	760,558	760,558	760,558	760,558
Adjusted R-squared	0.638	0.660	0.657	0.672	0.579	0.611	0.603	0.622

Note: Table presents the coefficients from a regression where the outcome variables are the standardized math and reading test scores. All specifications include threshold dummies indicating whether a student's score was within half a standard deviation of the teacher's judgment of student achievement level. Lagged Teacher Private Information (TPI<sub>t-1</sub>) is the difference between a student's lagged standardized score and the median score of the achievement level corresponding to the previous year's teacher's judgment on the student's achievement in Math (columns (1)-(4)) and Reading (columns (5)-(8)). The baseline ethnicity is white. Other ethnicities are Asian, Native American, and multiethnic. Additional controls include any pre-existing scores for the students across both subjects, in addition to dummy variables for repeated grade and repeated test. Classroom and school characteristics include all of the characteristics for the students pooled at the classroom, school-year, and school-grade levels. All specifications account for teacher, year, and grade fixed effects. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table A8: Regression of VA Estimates (Excluding  $TPI_{j',t-1}$ ) on Average Lagged Teacher Private Information

	Math	Reading
Outcome Var:	VA w/o TPI	VA w/o TPI
Average TPI at class level	-0.035*** (0.002)	-0.012*** (0.001)
Constant	0.010*** (0.001)	0.003*** (0.000)
Observations	29,632	29,632
R-squared	0.009	0.006

Notes: Table A8 presents the coefficients of the OLS regression where the outcome variable is the teacher VA estimates that are estimated without including the lagged Teacher Private Information ( $TPI_{j',t-1}$ ) on the average  $TPI_{j',t-1}$  for that specific year. All specifications use class size as weight. The sample of teachers does not include the whole sample as depicted in A1, as the observations are limited to those that have VA estimates attached to them. Nonetheless, the sample characteristics are substantially similar. For example, the ratio of female teachers in the whole sample is %89.8, whereas the ratio of female teachers with VA estimates is %89.5. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A9: Forecast Bias in VA Estimates when Omitting  $TPI_{i,j',t-1}$

	Math	Reading
Teacher VA	-0.011*** (0.001)	-0.031*** (0.002)
Observations	510,772	510,841
R-squared	0.001	0.001

Notes: To estimate the forecast bias, we implemented the following steps for each subject: 1) regress  $TPI_{i,j',t-1}$  on the same set of controls as in eq.(1) except for TPI (see Table 2) and recover the residuals ( $r1$ ), 2) regress the residuals from the first-stage of VA estimation (without TPI) on  $r1$ , 3) calculate predicted values from that regression ( $\rho*r1$ ), 4) regress  $\rho*r1$  on value-added estimates that do not include TPI in the first stage. The number of observations is reduced due to not having VA estimates for teachers without previous year information. Bootstrapped standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## B How Measurement Error/Omitted Teacher Private Information Impact Value-Added Estimates

In this appendix, we show how the usual two-step value-added (VA) estimator is biased whenever (i) the observed prior test score differs from the true prior score by a component of teacher-private information, and (ii) that component is omitted. Importantly, we do *not* assume purely classical measurement error; rather, we explicitly decompose the measurement gap into our measure of teacher private information plus an idiosyncratic noise term.

### B.1 Simplified Underlying Production Technology

Let  $Y_{it}$  denote student  $i$ 's test score in year  $t$ , and let  $Y_{i,t-1}^*$  be that student's true prior achievement. The true production function is

$$Y_{it} = \beta Y_{i,t-1}^* + \phi_{j(i,t)} + \varepsilon_{it}, \quad E[\varepsilon_{it} \mid Y^*, j] = 0, \quad (1)$$

where  $\beta$  captures persistence,  $\phi_{j(i,t)}$  is the true teacher effect, and  $\varepsilon_{it}$  is mean-zero conditional on  $Y^*$  and teacher assignment.

### B.2 Decomposition of the Measurement Gap

We only observe the noisy lagged score  $Y_{i,t-1}$ . Instead of classical error, write

$$Y_{i,t-1}^* = Y_{i,t-1} + m_{i,t-1}, \quad m_{i,t-1} = \delta TPI_{j'(i,t-1)} - \eta_{i,t-1}, \quad (2)$$

so that

$$Y_{i,t-1}^* - Y_{i,t-1} = \delta TPI_{j'(i,t-1)} - \eta_{i,t-1}. \quad (3)$$

Here  $TPI_{j'}$  is the teacher-private component (mean-zero across students but varying by teacher) and  $\eta$  is idiosyncratic noise with  $E[\eta \mid j] = 0$ .

## Step 1: First-Stage Regression

The usual first-stage is

$$Y_{it} = \lambda Y_{i,t-1} + u_{it}. \quad (4)$$

Substitute (1) and (2):

$$\begin{aligned} Y_{it} &= \beta \left( Y_{i,t-1} + \delta TPI_{j'} - \eta_{i,t-1} \right) + \phi_{j(i,t)} + \varepsilon_{it}, \\ u_{it} = Y_{it} - \lambda Y_{i,t-1} &= (\beta - \lambda) Y_{i,t-1} + \underbrace{\beta \delta TPI_{j'(i,t-1)}}_{\text{sorting bias}} - \underbrace{\beta \eta_{i,t-1}}_{\text{noise inflation}} + \phi_{j(i,t)} + \varepsilon_{it}. \end{aligned}$$

Thus, the first-stage residual is

$$u_{it} = (\beta - \lambda) Y_{i,t-1} + \beta \delta TPI_{j'(i,t-1)} - \beta \eta_{i,t-1} + \phi_{j(i,t)} + \varepsilon_{it}. \quad (5)$$

## Step 2: Recovering Teacher Value-Added

For expositional reasons, we simplified this step and regresses  $u_{it}$  on teacher indicators:

$$u_{it} = \hat{\phi}_{j(i,t)} + e_{it}. \quad (6)$$

Taking expectations conditional on teacher  $j$ , and using  $E[\eta | j] = 0$  but allowing  $E[TPI | j] \neq 0$ , gives

$$E[\hat{\phi}_j] = \phi_j + \underbrace{\beta \delta E[TPI_{j'} | j]}_{\text{sorting bias}} + \underbrace{(\beta - \lambda) E[Y_{i,t-1} | j]}_{\text{attenuation bias}}. \quad (7)$$

Hence omitting  $TPI$  imparts a bias of  $+\beta \delta E[TPI | j]$ , in addition to the usual attenuation  $(\beta - \lambda)E[Y_{i,t-1} | j]$  when  $\lambda < \beta$ .

By decomposing measurement error into a teacher-private component and classical noise, we see explicitly how omitting  $TPI$  both (i) induces a sorting-bias term of magnitude  $\beta \delta E[TPI | j]$ , and (ii) leaves the usual attenuation-type bias from mis-measuring the lagged score.