CEDR
Center for Education
Data & Research

Uneven Playing Field? Assessing the Inequity of Teacher Characteristics and Measured Performance Across Students*

Dan Goldhaber, University of Washington Bothell

Lesley Lavery, Macalaster College

Roddy Theobald, University of Washington Bothell

Abstract: Policymakers aiming to close the well-documented achievement gap between advantaged and disadvantaged students have increasingly turned their attention to issues of teacher quality. A number of studies have demonstrated that teachers are inequitably distributed across student subgroups by input measures like experience and qualifications, as well as output measures like value added estimates of teacher performance, but these tend to focus on either individual measures of teacher quality or particular school districts. In this study, we present a comprehensive, descriptive analysis of the inequitable distribution of both input and output measures of teacher quality across various indicators of student disadvantage across all school districts in Washington State. We demonstrate that in elementary, middle school, and high school classrooms, virtually every measure of teacher quality we examine—experience, licensure exam scores, and value-added—is inequitably distributed across every indicator of student disadvantage—free/reduced lunch status, underrepresented minority, and low prior academic performance. Finally, we decompose these inequities to the district, school, and classroom level, and find that patterns in teacher sorting at all three levels contribute to the overall teacher quality gaps.

Suggested citation:

Goldhaber, D., Lesley, L., and Theobald, R. (2014). Uneven Playing Field? Assessing the Inequity of Teacher Characteristics and Measured Performance Across Students. CEDR Working Paper 2014-4. University of Washington, Seattle, WA.

You can access other CEDR publications at
http://www.CEDR.us/publications.html

Follow us on twitter @CEDR_US

I.   Introduction

State and federal policymakers have actively sought to close achievement gaps between advantaged and disadvantaged students through a variety of mechanisms. While many factors contribute to measurable gaps in student performance, policymakers have increasingly turned their attention to issues of teacher quality. The focus on teachers is driven by a growing body of work that shows teacher quality to be the most important *schooling* factor in predicting academic success (Chetty et al., 2011; Rivkin et al., 2005; Rockoff, 2004) as well as evidence that various teacher characteristics, such as a teacher's classroom experience or value added measure of effectiveness, are distributed inequitably across student subgroups.

This paper provides the first comprehensive descriptive analysis of the inequitable distribution of both input (e.g., experience and credentials) and output (e.g., estimates of performance) measures of teacher quality across indicators of student disadvantage for a single state. We demonstrate that in Washington elementary, middle school, and high school classrooms, virtually every measure of teacher quality—experience, licensure exam score, and value-added estimates of effectiveness—is inequitably distributed across *every* indicator of student disadvantage—free/reduced lunch status, underrepresented minority, and low prior academic performance (the sole exception being licensure exam scores in high school math classrooms). For each combination of teacher quality measure and student disadvantage indicator, we calculate the difference between advantaged and disadvantaged students in exposure rates to less-qualified teachers (the "teacher quality gap"), and decompose this gap to the district, school, and classroom levels. We generally (but not always) find that most inequity comes from teacher sorting across districts and schools as opposed to sorting of teachers across classrooms in schools, but that patterns in teacher sorting at all three levels contribute to the overall teacher quality gaps.

The paper proceeds as follows. In the next section we review previous work on the inequitable distribution of teacher quality across student subgroups. We then describe the Washington State dataset we employ for this study. Finally, we present our methods and results and conclude with a discussion of policy implications.

II.   Background

A sizeable body of literature documents considerable inequities in the distribution of teacher quality (a term we use generically to refer to both input and output measures of teacher qualifications). This is not a new finding for teacher input measures (e.g., experience and licensure test scores). More than a decade ago, Lankford and colleagues (2002) used the New York state education workforce database to examine the distribution of teacher qualifications (teacher experience, degree, certification and college of attendance) throughout the state. They first examine all teachers in the state, and then decompose their analysis to see how quality varies between regions and labor markets. Focusing on the 10th, 50th and 90th percentiles of the distributions of these measures of teacher quality, they find that low-qualified teachers in New York are much more likely to teach in schools with higher proportions of poor, minority and low-performing students, particularly in urban areas.

Clotfelter, Ladd and Vigdor (2005) rely on micro-level data from North Carolina to examine the distribution of teacher experience for black and white students. They find that black students are much more likely to be in a classroom with a novice teacher than their white student peers (e.g., black 7th graders are 54 percent more likely to have a novice teacher in math and 38 percent more likely to have a novice teacher in English than white students). The authors decompose these differences into district, school, and classroom effects, and find considerable

effects at each level: in math, for instance, 38% of the gap is due to teacher sorting across district, 37% is due to teacher sorting across schools within districts, and 25% is due to teacher sorting across classrooms within schools.[1]

Two recent papers by Kalogrides and colleagues build on this prior work by focusing exclusively on student and teacher sorting within schools. Kalogrides and Loeb (2013) link student and teacher data from three large urban school districts to examine teacher sorting and find differences in achievement, racial, and socioeconomic composition of classrooms within schools. Consistent with the findings in Clotfelter et al. (2005), classrooms with the highest composition of high-need students (low-achieving, poor, and minority students) were most likely to have a novice teacher. Kalogrides, Loeb and Beteille (2013) further examine the extent to which teacher sorting occurs within schools using data from just one of the urban districts used in prior analyses to focus more explicitly on initial assignment. They find that less experienced, minority and female teachers are initially placed with lower achieving students than their more experienced, white, male peers.

Until recently, the most widely available proxies for teacher quality have been input variables like teacher credentials or experience. While the studies cited above, relying on these proxies, suggest that teacher quality is inequitably distributed across student subgroups, credentials and experience may be only weakly correlated with a teacher's contribution to student achievement (Aaronson et al., 2007; Goldhaber, 2008; Goldhaber and Brewer, 2000; Hanushek, 1997; Rivkin et al., 2005). In light of this, scholars have begun to explore how teacher effectiveness, as estimated by value added models, is distributed across student subgroups. Sass and colleagues (2010) use student-level data from Florida and North Carolina to compare teacher

---

[1] A follow up paper by the same authors (2006) explores the sorting of teachers and principals to high- and low-poverty schools.

value-added in high-poverty (>70% free and reduced price lunch (FRL) students) and lower-poverty (<70% FRL students) elementary schools. They find that teachers in high-poverty schools tend to have lower value-added than those in other schools though the magnitude of this finding is small and inconsistent across contexts. These differences are largely driven by the higher concentration of ineffective teachers in high-poverty schools (teachers at the top of the effectiveness distribution are similarly distributed across school settings). Similarly, Glazerman and Max (2012) find that, in a sample of 10 selected school districts in seven states, low-income students have unequal access to the highest-performing teachers at the middle school but not elementary school level. The authors find variation in the distribution of teacher performance within and among the districts studied.

Most recently, Isenberg and colleagues (Isenberg, Max, Gleason, Potamites, Santillano, Hock, and Hansen, 2013) explore the distribution of teacher effectiveness across 29 diverse school districts. They define a district's "effective teaching gap" as the difference in average value-added between teachers of advantaged students (those eligible for a free or reduced-price lunch) and teachers of their more advantaged (non-FRL) peers, and find consistent and significant effective teaching gaps. These results differ little across time (they analyze data from the 2008-09 through 2010-11 school year), though some districts have a smaller effective teaching gap than others. These gaps also persist under several sensitivity analyses (one controlling for the distribution of effectiveness across racial and ethnic subgroups). The authors conclude that effective teaching gaps in districts are due more to teacher assignment to schools than to teacher assignment to classrooms within schools.

Taken together, these studies clearly indicate that teacher quality is inequitably distributed across indicators of student disadvantage—regardless of the definition of teacher

quality and student disadvantage—in predictable ways. Table 1 summarizes the combinations of teacher quality measures, student disadvantage indicators, and grade levels that have been discussed in the existing literature.

In this paper, we utilize data from Washington State to quantify the inequitable distribution of teacher quality across student subgroups for *each* combination of school level, teacher quality variable, and student disadvantage category in Table 1. In doing so, we aim to make three distinct contributions to the existing literature. First, we provide the first comprehensive analysis of the inequitable distribution of both input (experience and credentials) and output (effectiveness) measures of teacher quality across different indicators of student disadvantage (family income, race, and prior achievement) using data from a single state.

Second, we decompose these "teacher quality gaps" into district, school, and classroom effects. To our knowledge, only one prior paper (Clotfelter et al., 2005) has done this, and in this paper the authors report estimates for only one teacher characteristic (experience) across one indicator of student disadvantage (minority) at one school level (secondary).[2] Our results provide a broader understanding of the degree to which, at least in one state, inequity is explained by teacher sorting across districts, across schools within a district, and across classrooms within a school.

Finally, following Sass et al. (2010), we focus on the lower tail of the teacher quality distribution (i.e., the probability of getting a very poor teacher) to investigate whether average differences in teacher quality (e.g., as reported in Isenberg et al., 2013) may mask inequities in exposure to teachers at the bottom end of the quality distribution. We build on the Sass paper by

---

[2] Isenberg et al. (2013) decompose their effective teaching gap into school and classroom effects, but cannot consider cross-district sorting.

considering our full range of student disadvantage indicators, rather than just student poverty level, as well as a full range of teacher quality measures, rather than just value-added.

III.   Data

The data for this study are derived primarily from four administrative databases maintained by Washington State's Office of Superintendent of Public Instruction (OSPI): the Comprehensive Education Data and Research System (CEDARS), the Student Testing Database, the Washington State S-275 personnel report, and the Washington State Credentials database. We use these databases to create a longitudinal dataset linking students to standardized test scores and their teachers in math and reading courses in grades 3-10 in the 2011-12 school year.[3] Our analysis focuses on three student variables and three teacher variables, each of which we discuss below.

Student variables: CEDARS and Student Tests

The CEDARS database, maintained by OSPI and designed to provide longitudinal data linking student and teacher schedules, includes an indicator for whether each student in the state is eligible for free or reduced-price lunch (FRL). This database also tracks the race and ethnicity of each student in the state. We create an indicator for "underrepresented minority" students— American Indian, black, and Hispanic—and use this and the FRL measure as two indicators of student disadvantage.

The state's Student Testing Database includes student test scores on the MSP, an annual state assessment of math and reading given to students in grades 3 through 8. This allows us to

---

[3] Due to a data-reporting error, our dataset does not include any elementary school students from the Tacoma School District

observe a prior year test score in reading and math for each student in grades 4-9 who was enrolled in Washington State schools the prior year and took the state exam. In addition to using these scores to calculate value-added estimates of teacher effectiveness, we create an indicator for whether each student scored in the lowest quartile of the test in the prior tested grade and year and use this indicator as a third measure of student disadvantage.

The Student Testing Database also contains two types of high school test scores. All 10th grade students in Washington State take the High School Proficiency Exam (HSPE) in reading, but students in grades 9 and 10 take different End-Of-Course (exams) in math depending on the math course they are enrolled in: either algebra or geometry. We use these test scores to calculate value-added measures of teacher performance in high school, discussed below.

Teacher input measures: S-275 and credentials database.

The S-275 database contains information from OSPI's personnel-reporting process, and includes a record of all certified employees in school districts as well as a measure of each employee's teaching experience in the state. Like many researchers (Anzia and Moe forthcoming; Clotfelter et al. 2005; Kalogrides and Loeb 2013; Koski and Horng 2007), we use these detailed data to create an indicator for "novice teachers" with two or fewer years of experience.

The Washington State Credentials database contains information on the licensure/certification status of all teachers in Washington, including when and where teachers obtained their initial teaching certificates. This database also includes teachers' test scores on the Washington Educator Skills Test – Basic, or WEST-B, a standardized test that all teachers must pass prior to entering a teaching training program. We calculate the average WEST-B score

across math, reading, and writing from the *first* time each teacher took the test.[4] For each teacher linked to WEST-B scores (generally teachers who entered the workforce after August 2002), we create an indicator for whether the teacher falls in the lowest 10% of the distribution of all average test scores.

Teacher output measures: prior value-added measures of teacher effectiveness

A growing body of literature uses value-added models (VAMs) to identify the contribution that individual teachers make toward student learning gains (e.g. Aaronson et al. 2007; Goldhaber and Hansen 2010; McCaffrey et al. 2004, 2009). The goal of these VAMs is to isolate the impact of individual teachers on student achievement from other factors (such as family background or class size) that influence achievement. The value-added estimate for teacher *j* in subject *s* in year *t* is calculated from the following VAM:[5]

$$Y_{ijst} = \beta_0 + \beta_1 Y_{i(t-1)} + \beta_2 X_{it} + \tau_{jst} + \varepsilon_{ijst}$$

Y_{ijst} is the state test score for each student *i* with teacher *j* in subject *s* (math or reading) and year *t,* normalized within grade and year; $Y_{i(t-1)}$ is a vector of the student's scores the previous year in both math and reading, also normalized within grade and year; X_{it} is a vector of student attributes in year *t* (gender, race, eligibility for free/reduced price lunch, English language learner status, gifted status, special education status, learning disability status); and τ_{jst} is a fixed effect that captures the contribution of teacher *j* to student test scores in subject *s* and year *t*. We adjust all

---

[4] Teachers may take the test multiple times to get a passing score on all three tests, so we use the first score to ensure that scores are comparable across teachers.

[5] The base model is the same as the model estimated by Isenberg et al. (2013). We make slight modifications to this model to estimate VAMs in high school. For math, we only estimate the model for 9th grade students enrolled in Algebra who took the Algebra EOC exam at the end of the year (using 8th scores as prior year test scores). For reading, the dependent variable is the student's HSPE score in 10th grade. However, students are not tested in reading in 9th grade, so the prior year test scores are the student's 8th grade test scores. We then include two teacher fixed effects—one for the 9th grade reading teacher and one for the 10th grade reading teacher—to account for combined contributions to the student's 10th grade test score.

teacher effect estimates using empirical Bayes (EB) methods.[6] For each student, we use each

teacher's VAM estimate from the *prior* school year (when the student was not in the teacher's

class)[7], and create indicators for whether the teacher's value added falls in the lowest decile of

the distribution of all value added estimates in the state.

IV. Methods

Let D_{ijkl} be an indicator of disadvantage (FRL, URM, or low prior performance) for

student i in classroom j within school k and district l ($D_{ijkl} = 1$ if the student is disadvantaged and

$D_{ijkl} = 0$ otherwise). Likewise, let X_{ijkl} be an indicator of low quality (novice, low credential exam

score, or low prior VAM estimate) for the teacher of student i in classroom j within school k and

district l ($X_{ijkl} = 1$ if the student's teacher has low quality and $X_{ijkl} = 0$ otherwise). For each

combination of student disadvantage indicator and teacher low quality indicator, we can

calculate the "exposure rate" of disadvantaged students to low-quality teachers via the following

exposure equation:

$$E_D\left(X_{ijkl}\right) = \frac{\sum_i \sum_j \sum_k \sum_l D_{ijkl} X_{ijkl}}{\sum_i \sum_j \sum_k \sum_l D_{ijkl}}$$

The numerator of $E_D\left(X_{ijkl}\right)$ is the total number of disadvantaged students who have a

low-quality teacher (summed over students, teachers, schools, and districts), while the

---

[6] The standard empirical Bayes method shrinks estimates back to the grand mean of the population. Note, however, that standard empirical Bayes adjustment does not properly account for the uncertainty in the grand mean, suggesting the estimates are shrunk too much (McCaffrey et al., 2009). But recent evidence (Herrmann et al., 2013) also suggests that shrinkage improves the estimates for teachers "hard-to-predict" students. We use the standard approach that's been commonly estimated in the literature (an appendix on empirical Bayes shrinkage is available from the authors by request).

[7] We use an estimate of each teacher's *prior* performance so that this measure of teacher quality—like experience and licensure scores—is measurable when students are assigned to classes (see Kalogrides and Loeb, 2013).

denominator is the total number of disadvantaged students. Thus $E_D\left(X_{ijkl}\right)$ is simply the percent

of disadvantaged students who are assigned to a low-quality teacher. We can also calculate the

equivalent exposure rate for non-disadvantaged students:

$$E_{ND}\left(X_{ijkl}\right) = \frac{\sum_i \sum_j \sum_k \sum_l (1 - D_{ijkl}) X_{ijkl}}{\sum_i \sum_j \sum_k \sum_l (1 - D_{ijkl})}$$

For each combination of student disadvantage indicator and teacher low quality indicator,

then, we define the overall "teacher quality gap" as the difference in exposure rates to low-

quality teachers between disadvantaged students and non-disadvantaged students:

$$Gap^{overall} \equiv E_D\left(X_{ijkl}\right) - E_{ND}\left(X_{ijkl}\right)$$

The teacher quality gap gives a snapshot of the inequitable distribution of teacher quality

across students in the state: a positive value indicates that disadvantaged students are more likely

to be assigned to a low-quality teacher, while a negative value means they are less likely.

However, this teacher quality gap (if it exists) can arise from three sources: teacher

sorting across districts (e.g., low-quality teachers may be more likely to teach in districts with

more disadvantaged students); teacher sorting across schools within districts (e.g., within

districts, low-quality teachers may be more likely to teach in schools with more disadvantaged

students); and/or teacher sorting across classrooms within schools (e.g., within schools, low-

quality teachers may be more likely to teach in classrooms with more disadvantaged students).

Therefore, following Clotfelter et al. (2005), we decompose the teacher quality gap into terms

related to district-level sorting, school-level sorting, and classroom-level sorting. We first

calculate the average exposure rates to low-quality teachers within each district l and school k (n_l

is the number of students in the district and n_{kl} is the number of students in the school).

$$\overline{X}_l = \frac{1}{n_l}\sum_i \sum_j \sum_k X_{ijkl} \text{ and } \overline{X}_{kl} = \frac{1}{n_{kl}}\sum_i \sum_j X_{ijkl}$$

We can then use these terms to decompose the overall teacher quality gap $Gap^{overall}$ into three terms:

$$
\begin{aligned}
Gap^{overall} &\equiv E_D\left(X_{ijkl}\right) - E_{ND}\left(X_{ijkl}\right) \\
&= \left[E_D\left(\overline{X}_l\right) - E_{ND}\left(\overline{X}_l\right)\right] \\
&\quad + \left[\left(E_D\left(\overline{X}_{kl}\right) - E_{ND}\left(\overline{X}_{kl}\right)\right) - \left(E_D\left(\overline{X}_l\right) - E_{ND}\left(\overline{X}_l\right)\right)\right] \\
&\quad + \left[\left(E_D\left(X_{ijkl}\right) - E_{ND}\left(X_{ijkl}\right)\right) - \left(E_D\left(\overline{X}_{kl}\right) - E_{ND}\left(\overline{X}_{kl}\right)\right)\right] \\
&\equiv Gap^{district} + Gap^{school} + Gap^{class}
\end{aligned}
$$

The first term, $Gap^{disttrict} \equiv \left[E_D\left(\overline{X}_l\right) - E_{ND}\left(\overline{X}_l\right)\right]$, is the "district effect" (following Clotfelter et al., 2005), and can be interpreted as the average difference in *district-level* exposure rates to low-quality teachers between disadvantaged and non-disadvantaged students. If this value is positive, it means that disadvantaged students are more likely to attend districts with high percentages of low-quality teachers.

The second term, $Gap^{school} \equiv \left[\left(E_D\left(\overline{X}_{kl}\right) - E_{ND}\left(\overline{X}_{kl}\right)\right) - \left(E_D\left(\overline{X}_l\right) - E_{ND}\left(\overline{X}_l\right)\right)\right]$, is the "school effect", or the difference in average school-level exposure rates to low-quality teachers between disadvantaged and non-disadvantaged students *subtracting out the difference in average district-level exposure rates*. Gap^{school} can be re-written as $\left[\left(E_D\left(\overline{X}_{kl}\right) - E_D\left(\overline{X}_l\right)\right) - \left(E_{ND}\left(\overline{X}_{kl}\right) - E_{ND}\left(\overline{X}_l\right)\right)\right]$, which demonstrates that the school effect is also the difference in school-level rates of low-quality teachers between disadvantaged and non-disadvantaged students *relative to the percent of*

low-qualified teachers in those students' districts. Thus a positive school effect means that disadvantaged students are more likely to attend schools with a higher percentage of low-quality teachers than non-disadvantaged students within the same district.

The last term, $Gap^{class} \equiv \left[\left(E_D \left(X_{ijkl} \right) - E_{ND} \left(X_{ijkl} \right) \right) - \left(E_D \left(\overline{X}_{kl} \right) - E_{ND} \left(\overline{X}_{kl} \right) \right) \right]$, is the "classroom effect", which simply subtracts the sum of the school and district effects from the overall teacher quality gap. Gap^{class} can be re-written as

$\left[\left(E_D \left(X_{ijkl} \right) - E_D \left(\overline{X}_{kl} \right) \right) - \left(E_{ND} \left(X_{ijkl} \right) - E_{ND} \left(\overline{X}_{kl} \right) \right) \right]$, which demonstrates that the classroom effect is also the difference in exposure rates to low-quality teachers between disadvantaged and non-disadvantaged students *relative to the percent of low-quality teachers in those students' schools*. Thus a positive classroom effect means that disadvantaged students are more likely to be assigned a low-quality teacher than non-disadvantaged students within the same school.

V. Results

We present our results in two steps. First, to clarify our methods and take a close look at the inequitable distribution of one teacher characteristic across students in one grade level, we focus solely on 4th grade classrooms and investigate the distribution of novice teachers across indicators of student disadvantage. Then, we repeat this procedure for all three indicators of teacher quality (experience, credential exam scores, and value-added) and representative grades for all three school levels (elementary, middle school, and high school).

Distribution of novice teachers in 4th grade classrooms

Table 2 gives an overview of the distribution of novice teachers across all three indicators of student disadvantage—FRL (free/reduced lunch eligibility), URM (underrepresented minority), and low prior performance (lower quartile prior year test scores)—for 4th grade classrooms in Washington State.[8] The first row of results gives the exposure rates for disadvantaged and non-disadvantaged students ($E_D(X_{ijkl})$ and $E_{ND}(X_{ijkl})$ from section IV, respectively) for each indicator of disadvantage, as well as the "teacher quality gap" ($Gap^{overall}$ from section IV). We can see that for each indicator of disadvantage, but particularly for URM students, disadvantaged 4th-grade students are more likely to be assigned to a novice teacher than non-disadvantaged 4th-grade students (and each teacher quality gap is statistically significant at the .05-level).

Some interesting patterns emerge when we decompose these teacher quality gaps into district, school, and classroom effects (shown in the Panel 1 of Table 2). The effects themselves are in the "Gap" column in Table 2, while the terms defined in section IV and used to calculate these effects— $E_D(\overline{X}_l)$ and $E_{ND}(\overline{X}_l)$ for the district effect, $\left(E_D(\overline{X}_{kl}) - E_D(\overline{X}_l)\right)$ and $\left(E_{ND}(\overline{X}_{kl}) - E_{ND}(\overline{X}_l)\right)$ for the school effect, and $\left(E_D(X_{ijkl}) - E_D(\overline{X}_{kl})\right)$ and $\left(E_{ND}(X_{ijkl}) - E_{ND}(\overline{X}_{kl})\right)$ for the classroom effect—are in the other columns. Across each indicator of disadvantage, the school and district effects are larger than the classroom effects (and are statistically significant[9]), but the relative magnitudes vary depending on the definition of student disadvantage. For example, the teacher quality gap for FRL students appears to be driven equally by teacher sorting across districts and teacher sorting across schools within a district. On the

---

[8] We choose to focus on the distribution of early-career teachers because is well known that teachers become more productive early in their careers (e.g., Rice 2013).
[9] We test the null hypothesis that each effect equals zero using a two-sided t-test.

other hand, the teacher quality gap for URM students appears to be driven primarily by teacher sorting across districts; i.e., URM students are much more likely to attend a district with a high percentage of novice teachers than non-URM students. In none of the three cases do we see evidence that student sorting across classrooms within schools contributes significantly to the teacher quality gap.

We report the teacher quality gap and district, school, and classroom effects for each combination of school level, indicator of student disadvantage, and indicator of low teacher quality in the next sub-sections. But before we proceed, we dig a little deeper into the inequitable distribution of novice teachers in 4th grade. First, **Figure 1** shows the observed distribution of teacher experience in 4th grade classrooms by student FRL status (the green vertical line indicates our cutoff for "novice teachers", while the other vertical lines indicate the means for each group). We see that distribution of teacher experience for FRL 4th-grade students is weighted more heavily towards inexperienced teachers, and the average teacher experience for FRL 4th-grade students is almost a full year less than the average teacher experience for non-FRL 4th grade students.[10]

Next, **Figure 2** plots the exposure rate to novice teachers for 4th-grade FRL students against the exposure rate for 4th-grade FRL students *within* the 23 largest districts in the state. While the majority of districts fall above the 45-degree line—indicating the 4th-grade FRL students in these districts are more likely to be assigned a novice teacher than 4th-grade non-FRL students—there are a number of districts below the 45-degree line. In other words, there is some variation across districts in terms of the inequitable distribution of novice teachers across FRL

---

[10] As we argue in Section III, we believe that differences in exposure rates to low-qualified teachers may be more important than the difference in mean teacher characteristics between disadvantaged and non-disadvantaged students. However, we replicate all our analyses using mean teacher characteristics, and find similar patterns. These supplemental results are available from the authors upon request.

and non-FRL students.

Finally, the last two panels of Table 2 explore whether the teacher quality gap is higher in some types of districts than others. Panel 2 shows that the distribution of novice teachers across both FRL and URM students is most inequitable within the most disadvantaged districts. On the other hand, Panel 3 shows that the distribution of novice teachers across each of the student disadvantage indicators is most inequitable in the smallest districts. Again, this simply demonstrates that the magnitude (and even direction) of the inequitable distribution of novice teachers across indicators of student disadvantage varies across districts.

Distribution of low-qualified teachers across all student indicators and grade levels

The inequitable distribution of teacher experience is already well-established in the literature (e.g., Clotflelter et al., 2005; Lankford et al., 2002), but we now proceed to investigate teacher quality gaps for every combination of school level, student disadvantage indicator, and indicator of teacher quality. **Table 3** presents the overall teacher quality gap as well as the decompositions into district, school, and classroom effects.[11] The first row of results in Table 3 repeats the relevant results from Table 2 about the distribution of novice teachers across various indicators of student disadvantage in 4th grade. The remaining rows present the analogous results for other indicators of teacher quality—an indicator for whether the teacher fell into the lowest decile of value-added estimates the prior year ("Lowest decile prior VAM"), and an indicator for whether the teacher fell into the lowest decile of teacher credential exam scores ("Lowest decile WEST-B")—and other grade levels.

---

[11] We present results for 4th grade in elementary school, 7th grade math and reading in middle school, and 9th grade algebra and 10th grade reading in high school, but we also calculate results for other available grade levels and find consistent patterns. These results are available from the authors upon request.

We first focus on the teacher quality gap for each of these combinations, highlighted in bold in Table 3. Across *nearly every* combination of school level, student disadvantage indicator, and indicator of low teacher quality, the teacher quality gap is significant and positive; that is, disadvantaged students (regardless of definition) are more likely to have a low-quality teacher (regardless of definition) than non-disadvantaged students in the same grade level. The only exception is the distribution of teachers with low credential-exam scores across students in 9th-grade algebra classrooms, as none of these teacher quality gaps is statistically significant.

It is also interesting to note the variability in the magnitude of the teacher quality gaps in Table 3. The highest gap is for the distribution of teachers with low prior VAM estimates across students in 7th-grade math with low prior performance; 19.25% of low-performing 7th grade math students are assigned to a teacher with a low prior-year VAM estimate, compared to just 7.31% of higher performing math students in 7th grade (resulting in a teacher quality gap of 11.94%). A similarly large gap occurs for the same combination of student and teacher indicators in 7th-grade reading.[12] That said, large teacher quality gaps exist throughout Table 3, reinforcing the magnitude of the inequitable distribution of teacher quality across student subgroups in Washington State.

We next turn our attention to the decomposition of each of these teacher quality gaps. For most (but not all) combinations of school level, student disadvantage indicator, and indicator of low teacher quality, the largest effect is at the district level (i.e., disadvantaged students are more

---

[12] Isenberg et al. (2013) calculate the "effective teaching gap" as the difference in the *mean* value-added between advantaged and disadvantaged students. We replicate their procedure and find large differences at the means as well. For example, in both 7th grade math and reading, the effective teaching gap between students with low prior performance and students with not-low prior performance is .069 (i.e., the average low performing student has a 7th grade teacher whose performance is 7% of a standard deviation of student performance lower than the average not-low performing student). Further, we find that—in 7th grade—the majority of the effective teaching gap is attributable to the classroom level (unlike Isenberg et al. (2013) who find the majority of variation at the school level). Full results (calculated at the means) for each combination of school level, student disadvantage indicator, and measure of teacher quality are available from the authors upon request.

likely to attend districts with high percentages of low-qualified teachers than non-disadvantaged students in the same grade). For example, for nearly every teacher quality gap in 4th grade, the district effect explains over half of the teacher quality gap. There are interesting exceptions, however. For the two large teacher quality gaps in 7th grade discussed above, the majority of the gap can be explained by the classroom effect; in other words, within schools, 7th grade students with low prior performance are more likely to be assigned to *classrooms* with a teacher with low prior value-added estimates. This suggests that tracking within schools by prior performance may be a larger issue in middle schools than in the other school levels.

V.   Discussion and Conclusions

Our findings demonstrate that *every* measure of teacher quality—experience, licensure exam score, and value-added estimates of effectiveness—is inequitably distributed across *every* indicator of student disadvantage—free/reduced lunch status, underrepresented minority, and low prior academic performance—at virtually every school level in Washington State. We also demonstrate that patterns in teacher sorting across districts, across schools within districts, and across classrooms within schools *all contribute* to these teacher quality gaps. The teacher labor market literature provides a number of explanations for inequitable teacher sorting at each level.

Patterns in teacher retention, cross-district transfers, and hiring can influence the distribution of teacher quality across districts, and empirical evidence suggests that each process could contribute to the cross-district inequities we observe. A number of studies (e.g., Goldhaber et al., 2010; Guarino et al., 2006; Hanushek et al., 2004; Scafidi et al., 2007) have shown that teachers are more likely to leave districts with more disadvantaged students, meaning that these

districts need to hire more teachers each year. Yet prospective teachers are more likely to apply to districts with fewer disadvantaged students (Boyd et al., 2013; Engel et al., forthcoming), in part because teachers are generally paid using a single salary schedule that does not account for the difficulty of a teaching assignment, meaning that districts with more disadvantaged students also have fewer prospective teachers to choose from.[13] And new evidence (Goldhaber et al., 2014a) suggests that student teaching assignments could also contribute to these inequities: prospective teachers tend to do their student teaching at more advantaged schools, and these schools may use student teaching as a "screening process" to hire the most qualified prospective teachers.

Patterns in hiring and transfers may also contribute to the within-district, cross-school inequities we describe. The literature on within-district teacher transfers (e.g., Goldhaber et al., 2010; Scafidi et al., 2007) demonstrates that teachers are more likely to leave disadvantaged schools for another school in the district. This may be due to teacher's strong preferences for good working conditions, which are often correlated with the student demographics of a school (Horng, 2009; Ladd, 2011). Teacher collective bargaining agreements (CBAs) provide an opportunity for teachers to act on these preferences, since many CBAs contain provisions that protect senior teachers from involuntary transfers and grant senior teachers the right to voluntarily transfer to more desirable positions within the district. In a companion paper (Goldhaber et al., 2014c), we show that the probability that a teacher transfers out of a school with many disadvantaged students is particularly high in districts with strong CBA seniority transfer protections, suggesting that CBAs may contribute to within-district inequities.

---

[13] See Goldhaber et al. (2013b) for more discussion of the consequences of Washington State's single salary schedule.

Finally, there are a number of explanations for the within-school inequities we describe. There is evidence that principals reserve "favorable" classroom assignments for teachers with greater classroom success and higher exam licensure scores (Player, 2010), perhaps due to rigidities in teacher compensation structures. And in schools that "track" students by performance level, the inequities we observe (particularly at the middle school level) could be due in part to more qualified teachers being assigned to teach more "advanced" courses.

On the whole, it is not surprising that we observe large teacher quality gaps between advantaged and disadvantaged students given the evidence from this literature. But an emerging literature also suggests some potential solutions. For example, Clotfelter et al. (2008) find that a modest bonus to teachers who teach in high-poverty and low-performing schools in North Carolina decreased the mean teacher turnover rates in these schools by 17 percent. Grissom et al. (2014) find that, after a change in the involuntary transfer policy in Miami-Dade County Public Schools that gave administrators more flexibility in teacher assignments, principals of low-performing schools were able to identify low-performing teachers for transfer who would have been unlikely to leave on their own. So, while this paper documents the problem of teacher quality gaps in public schools, emerging evidence suggests that there may be solutions to close these gaps.

Bibliography

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*(1).

Anzia, S.F. & Moe, T.M. (Forthcoming). Collective bargaining, transfer rights, and disadvantaged schools. *Educational Evaluation and Policy Analysis.*

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, *31*(1), 83-117.

Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D.W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, *126*(4)*, 1593-1660.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J.L. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, 24(4), 377-392.

Clotfelter, C., Ladd, H. F., Vigdor, J., & Wheeler, J. (2006). High-poverty schools and the distribution of teachers and principals. *NCL Rev.*, 85, 1345.

Clotfelter, C., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, *92*(5), 1352-1370.

Engel, M., Jacob, B., & Curran, F. (forthcoming). New evidence on teacher labor supply. *American Educational Research Journal.*

Glazerman, S., & Max, J. (2011). Do low-income students have equal access to the Highest-performing teachers? (No. 6955). Mathematica Policy Research.

Goldhaber, D. (2008) Teachers matter, but effective teacher quality policies are elusive. In Ladd, H. and Fiske, E., ed., *Handbook of Research in Education Finance and Policy,* New York: Routledge, pp.146-165.

Goldhaber, D. & Brewer D.J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Education Evaluation and Policy Analysis, 22,* 129-45.

Goldhaber, D., Destler, K., & Player, D. (2007). Teacher labor markets and the perils of using hedonics to estimate compensating differentials in the public sector. (*SFRP Working Paper, No. 17*), Center for Reinventing Public Education.

Goldhaber, D., Gross, B., and Player, D. (2010). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management* 30(1), 57-87.

Goldhaber, D. & Hansen, M. (2010). Using performance on the job to inform teacher tenure decisions. (Brief 10). *National Center for Analysis of Longitudinal Data in Education Research*.

Goldhaber, D., Walch, J., & Gabele, B. (2013). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy*, *1*(1), 28–39.

Goldhaber, D., Krieg, J., Theobald, R., and Brown, N. (2014a). Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. Center for Education Data and Research.

Goldhaber, D., Krieg, J., Theobald, R., and Brown, N. (2014b). The line is shorter over here: The probability and time to first job of prospective teachers with different training. Center for Education Data and Research.

Goldhaber, D., Lavery, L., and Theobald, R. (2014c). Inconvenient truth? Do collective bargaining agreements help explain the movement of teachers within school districts? Center for Education Data and Research.

Grissom, J. A., Loeb, S., & Nakashima, N. A. (2014). Strategic involuntary teacher transfers and teacher performance: Examining equity and efficiency. *Journal of Policy Analysis and Management*, *33*(1), 112-140.

Guarino, C., Santibañez, L., & Daley, G.A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research, 76*(2), 173-208.

Hanushek, E. (1997) Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2)**,** 141-164.

Hanushek, E., Kain, J. & Rivkin, S. (2004). Why public schools lose teachers. *Journal of Human Resources* 39(2), 326-254.

Herrmann, M., Walsh, E., Isenberg, E., and Resch, A. (2013). Shrinkage of value-added estimates and characteristics of students with hard-to-predict achievement levels. Mathematica Policy Research.

Horng, E. L. (2009). Teacher tradeoffs: Disentangling teachers' preferences for working conditions and student demographics. *American Educational Research Journal*, *46*(3), 690-717.

Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., & Hock, H. (2013). Access to effective teaching for disadvantaged students. *National Center for Education Evaluation and Regional Assistance*, U.S. Department of Education.

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude and effects of student sorting within schools. *Educational Researcher*, 42(6), 304-316.

Kalogrides, D., Loeb, S., & Beteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, 86(2), 103-123.

Koski, W.S. & Horng, E. (2007) Facilitating the teacher quality gap? Collective bargaining agreements, teacher hiring and transfer rules, and teacher assignment among schools in California, *Education Finance and Policy, 262.*

Ladd, H. F. (2011). Teachers' Perceptions of Their Working Conditions How Predictive of Planned and Actual Teacher Movement?. *Educational Evaluation and Policy Analysis*, *33*(2), 235-261.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37-62.

McCaffrey, D.F., Lockwood, D., Louis, T.A., & Hamilton, L. (2004). *Journal of Educational and Behavioral Statistics*, *29*(1), 67-101.

Player, D. (2010). Nonmonetary compensation in the public teacher labor market. *Education Finance and Policy*, *5*(1), 82-103.

Rice, J.K. (2013). Learning from experience: Evidence on the impact and distribution of teacher experience and the implications for teacher policy. *Education Finance and Policy*, 8(3), 332-348.

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, *94*(2), 247-252.

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2010). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2), 104-122.

Scafidi, B., Sjoquist, D., and Stinebricker, T. (2007). Race, poverty, and teacher mobility. *Economics of Education Review* 26, 145-159.

Tables and Figures

Table 1. Summary of papers demonstrating inequitable distribution of teacher quality (rows) across indicators of student disadvantage (columns).

	Student FRL	Student minority	Student performance
All grades			
Teacher experience	Lankford et al. 2002 Sass et al. 2010 Clotfelter et al. 2006 Kalogrides & Loeb 2013	Lankford et al. 2002 Kalogrides & Loeb 2013	Kalogrides et al. 2013 Kalogrides & Loeb 2013
Teacher credentials	Lankford et al. 2002 Sass et al. 2010 Clotfelter et al. 2006	Lankford et al. 2002	
Teacher VAM	Isenberg et al. 2013		
Elementary Grades			
Teacher experience	Clotfelter et al. 2006 Sass et al. 2010 Kalogrides & Loeb 2013	Kalogrides & Loeb 2013	Lankford et al. 2002 Clotfelter et al. 2006 Kalogrides et al. 2013 Kalogrides & Loeb 2013
Teacher credentials	Clotfelter et al. 2006 Sass et al. 2010		Lankford et al. 2002 Clotfelter et al. 2006
Teacher VAM	Glazerman & Max 2011 Sass et al. 2010 Isenberg et al. 2013		
Secondary Grades			
Teacher experience	Clotfelter et al. 2005 Clotfelter et al. 2006 Kalogrides & Loeb 2013	Clotfelter et al. 2005 Kalogrides & Loeb 2013	Lankford et al. 2002 Clotfelter et al. 2005 Kalogrides et al. 2013
Teacher credentials	Clotfelter et al. 2006		Lankford et al. 2002
Teacher VAM	Glazerman & Max 2011 Isenberg et al. 2013		

Table 2. Overview of exposure rates to novice teachers in 4th grade classrooms by student disadvantage indicator and decomposition of differences

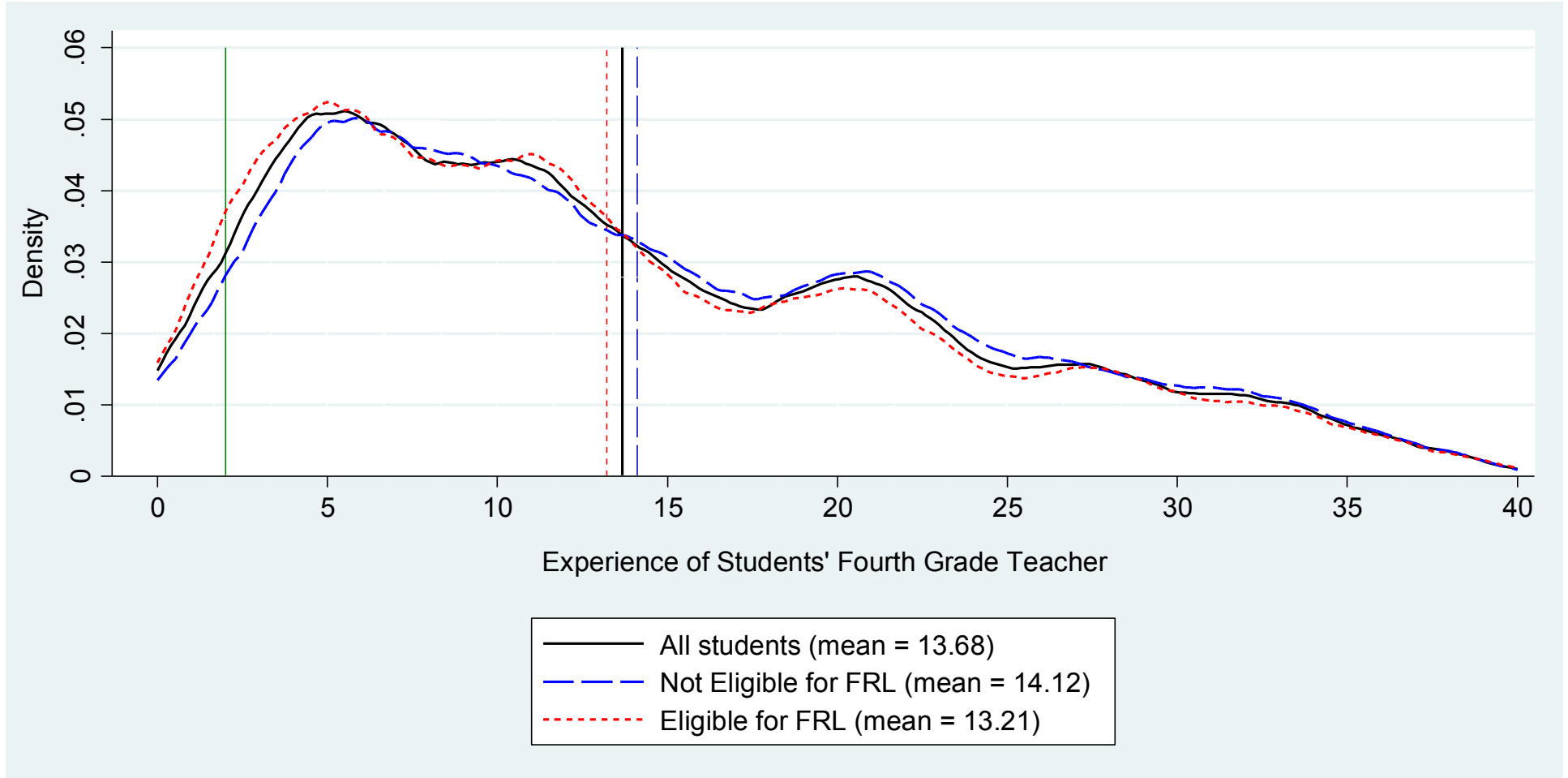| 4th Grade classrooms									
	By FRL Status			*By URM Status*			*Quintile of Prior Performance*		
Exposure to Novice Teacher (≤ 2 yrs exp)	**FRL**	**Non FRL**	**Gap**	**URM**	**Non URM**	**Gap**	**Lowest**	**Non Lowest**	**Gap**
State level	6.94%	5.54%	1.39%*	7.95%	5.64%	2.31%*	7.13%	5.92%	1.21%*
Panel 1: Decomposition of difference									
District Level	6.51%	5.95%	0.56%*	7.62%	5.75%	1.87%*	6.70%	6.07%	0.63%*
School Level	0.31%	-0.30%	0.61%*	0.28%	-0.09%	0.38%*	0.26%	-0.09%	0.34%*
Classroom Level	0.12%	-0.11%	0.22%	0.05%	-0.02%	0.06%	0.17%	-0.06%	0.24%
Panel 2: By quartile of district disadvantage (FRL, URM, or Low Performance)									
Lowest quartile (most advantaged)	5.69%	6.55%	-0.85%*	4.12%	4.32%	-0.21%	6.04%	6.25%	-0.22%
2nd quartile	4.09%	4.40%	-0.30%	4.94%	4.28%	0.66%	4.10%	3.21%	0.85%*
3rd quartile	7.34%	4.72%	2.62%*	5.10%	5.72%	-0.62%	7.07%	5.70%	1.37%*
Highest quartile (most disadvantaged)	9.09%	6.70%	2.40%*	11.67%	10.39%	1.28%*	9.86%	9.42%	0.44%
Panel 3: By quartile of district size									
Lowest quartile (smallest)	6.45%	3.05%	3.40%*	7.62%	4.01%	3.61%*	6.71%	4.24%	2.47%*
2nd quartile	3.98%	4.04%	-0.05%	5.02%	3.74%	1.29%*	4.32%	3.92%	0.40%
3rd quartile	9.11%	7.44%	1.67%*	9.97%	7.42%	2.55%*	10.14%	7.59%	2.55%*
Highest quartile (largest)	8.31%	7.36%	0.95%*	8.20%	7.68%	0.52%	7.19%	7.97%	-0.78%

Significance levels from two-sided t-test: *$p < 0.05$

Table 3. Exposure rates to low-quality teachers by grade level and student disadvantage indicator and decompositions of differences

| | Free/reduced priced lunch | | | | | | Underrepresented minority | | | | | | Low prior performance | | | | | |
| | By student FRL | | | *Decomposition of Difference* | | | By student URM | | | *Decomposition of Difference* | | | By Quintile of prior performance | | | *Decomposition of Difference* | | |
	FRL	Non FRL	**Gap**	District	School	Class	URP	Non URP	**Gap**	District	School	Class	Lowest	Non Lowest	**Gap**	District	School	Class
4th Grade																		
Novice (<2 years exp)	6.94%	5.54%	**1.39%***	0.56%*	0.61%*	0.22%	7.95%	5.64%	**2.31%***	1.87%*	0.38%*	0.06%	7.13%	5.92%	**1.21%***	0.63%*	0.34%*	0.24%
Lowest decile prior VAM	12.35%	8.41%	**3.93%***	1.83%*	1.29%*	0.81%*	12.34%	9.62%	**2.72%***	1.62%*	0.67%*	0.43%*	11.46%	9.91%	**1.55%***	1.45%*	0.02%	0.43%*
Lowest decile WEST-B	13.62%	10.07%	**3.54%***	2.45%*	0.78%	0.32%	13.03%	11.47%	**1.56%***	1.86%*	-0.64%	0.34%	12.83%	11.63%	**1.19%***	1.23%*	-0.18%	0.14%
7th Grade Math																		
Novice (<2 years exp)	9.18%	6.66%	**2.53%***	1.03%*	1.16%*	0.33%*	10.36%	7.03%	**3.33%***	1.71%*	1.22%*	0.40%*	9.70%	7.19%	**2.51%***	0.66%*	0.96%*	0.90*
Lowest decile prior VAM	13.59%	7.37%	**6.22%***	2.55%*	1.42%*	2.25%*	13.70%	9.02%	**4.68%***	2.33%*	0.80%*	1.55%*	19.25%	7.31%	**11.94%***	2.97%*	1.48%*	7.49%*
Lowest decile WEST-B	12.17%	8.05%	**4.11%***	2.60%*	0.66%*	0.85%*	14.05%	8.52%	**5.53%***	4.16%*	0.54%*	0.83%*	15.84%	7.95%	**7.89%***	1.35%*	3.15%*	3.39%*
7th Grade Reading																		
Novice (<2 years exp)	6.67%	4.85%	**1.82%***	-0.12%	1.06%*	0.88%*	7.27%	5.19%	**2.08%***	0.54%*	0.93%*	0.60%*	7.49%	5.08%	**2.35%***	-0.13%	0.84%*	1.64%*
Lowest decile prior VAM	12.30%	8.43%	**3.87%***	1.23%*	0.76%*	1.89%*	12.40%	9.44%	**2.96%***	0.43%*	0.84%*	1.69%*	17.79%	7.72%	**10.07%***	1.95%*	1.15%*	6.97%*
Lowest decile WEST-B	14.56%	7.14%	**7.42%***	6.68%*	0.34%	0.40%	15.09%	9.06%	**6.03%***	4.53%*	1.14%*	0.36%	15.24%	9.11%	**6.12%***	3.42%*	1.74%*	0.96%*
9th Grade Algebra																		
Novice (<2 years exp)	12.76%	9.23%	**3.53%***	2.20%*	1.14%*	0.18%	14.81%	9.59%	**5.23%***	4.58%*	0.80%*	-0.15%	13.02%	10.25%	**2.77%***	1.51%*	0.74%*	0.52%
Lowest decile prior VAM	12.72%	7.62%	**5.09%***	4.39%*	0.47%*	0.23%	14.77%	8.36%	**6.42%***	5.95%*	0.28%	0.19%	11.82%	9.40%	**2.42%***	1.94%*	0.25%	0.22%
Lowest decile WEST-B	11.18%	10.68%	**0.49%**	2.37%*	-1.34%*	-0.54%	10.32%	11.18%	**-0.86%**	0.42%	-1.19%*	-0.09%	11.09%	10.88%	**0.22%**	1.04%*	-0.65%*	-0.18%
10th Grade Reading																		
Novice (<2 years exp)	8.86%	7.38%	**1.48%***	0.95%*	0.50%*	0.03%	8.94%	7.69%	**1.25%***	0.92%*	0.44%*	-0.01%	8.81%	7.73%	**1.12%***	0.32%*	0.33%*	0.47%
Lowest decile prior VAM	11.32%	9.46%	**1.85%***	0.46%*	0.61%*	0.78%*	11.75%	9.73%	**2.01%***	0.64%*	0.36%*	1.01%*	12.63%	9.53%	**3.10%***	0.53%*	0.60%*	1.96%*
Lowest decile WEST-B	11.15%	8.09%	**3.05%***	2.50%*	-0.03%	0.58%*	11.44%	8.72%	**2.71%***	2.48%*	-0.17%	0.41%	10.95%	8.92%	**2.03%***	2.53%*	-1.08%*	0.58%

Significance levels from two-sided t-test: *p < 0.05

Figure 1. Observed distribution of teacher experience in 4th grade classrooms by student FRL status

Figure 2. Exposure rates to novice teachers in 4th grade classrooms by student FRL status for large districts