Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students

Dan Goldhaber[1], Lesley Lavery[2], and Roddy Theobald[1]

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 school, 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-price lunch status, underrepresented minority, and low prior academic performance. Finally, we decompose these inequities to the district, school, and classroom levels and find that patterns in teacher sorting at all three levels contribute to the overall teacher quality gaps.

Keywords: achievement gap; certification/licensure; descriptive analysis; economics of education; educational policy; educational reform; school/teacher effectiveness

State and federal policymakers have actively sought to close achievement gaps between advantaged and disadvantaged students through a variety of mechanisms. Although 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, Friedman, & Rockoff, 2013; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004) as well as evidence that various teacher characteristics, such as a teacher's classroom experience or estimated effectiveness, are distributed inequitably across student subgroups.

This article 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 state elementary school, 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-price lunch status (FRL), underrepresented minority (URM), 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.

[1]American Institutes for Research, Seattle, WA
[2]Macalester College, St. Paul, MN

Educational Researcher, Vol. XX No. X, pp. 1–15
DOI: 10.3102/0013189X15592622

The article 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 data set we employ, present our methods and results, and conclude with a discussion of policy implications.

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, Loeb, and Wyckoff (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 microlevel 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 are their White student peers (e.g., Black seventh graders are 54% more likely to have a novice teacher in math and 38% 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 districts, 37% is due to teacher sorting across schools within districts, and 25% is due to teacher sorting across classrooms within schools.[1]

Two recent articles 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 are 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. Although 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, Barrow, & Sander, 2007; Goldhaber, 2008; Goldhaber & 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, Hannaway, Xu, Figlio, and Feng (2010) use student-level data from Florida and North Carolina to compare teacher value added in high-poverty (>70% 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, although 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 (2011) 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 (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 disadvantaged students (those eligible for FRL) 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 school years 2008–2009 through 2010–2011), although some districts have a smaller effective teaching gap than others. These gaps also persist under several sensitivity analyses. 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.

There are a number of technical but important caveats to any analysis (such as the three studies discussed above as well as our own) that investigates the distribution of estimated teacher effectiveness across different types of students. Most importantly, different value-added models make different assumptions about how much of the differences in test scores between different types of students and schools should be attributed to teachers.[2] This means that—despite generally high correlations between estimates across different types of value-added models—the correlations between value-added estimates and student characteristics can vary depending on the model specification (Goldhaber, Walch, & Gabele, 2014). Importantly, though, the model specification utilized in the existing literature (and in our models) has the *least* correlation between the effectiveness estimates and student characteristics (Goldhaber, Walch, et al., 2014). Therefore, if anything, estimates of the teacher quality gap in value-added performance from the existing literature (and our article) are conservative.

In this article, 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 explored in the existing literature. In doing so, we aim to make three distinct

Table 1
Exposure Rates to Novice Teachers in Fourth-Grade Classrooms by Student Disadvantage Indicator and Decomposition of Differences (in percentages)

	By FRL Status			By URM Status			By Quintile of Prior Performance (Math and Reading Average)		
Exposure to Novice Teacher (≤2 years experience)	FRL	Non-FRL	Diff	URM	Non-URM	Diff	Lowest	Non-Lowest	Diff
State level	6.94	5.54	1.39*	7.95	5.64	2.31*	7.13	5.92	1.21*
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	–0.10	0.38*	0.26	–0.09	0.34*
Classroom level	0.11	–0.11	0.22	0.05	–0.02	0.06	0.17	–0.06	0.23
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	5.92	6.14	–0.22
2nd quartile	4.09	4.40	–0.30	4.94	4.28	0.66	4.06	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

Note. FRL = free or reduced-price lunch; URM = underrepresented minority; Diff = difference.
*$p < .05$ (two-sided t test).

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 article (Clotfelter et al., 2005) has done this, and in that article, the authors report estimates for only one teacher characteristic (experience) across one indicator of student disadvantage (minority) at one school level (secondary).[3] 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 et al. article by considering a 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.

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. Washington State is a prime setting for this investigation as it is close to the national median in terms of the percentage of students eligible for FRL (39.8%, or 28th out of 50 states and Washington, D.C.) and percentage of URM students (24.1%, or 29th out of 50 states and Washington, D.C.) in the state (U.S. Department of Education, 2010). We use the OSPI data to create a data set linking students to standardized test scores and their teachers in math and reading courses in Grades 3 through 10 in the 2011–2012 school year.[4] 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 FRL. This database also tracks the race and ethnicity of each student in the state. We create an indicator for URM 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 Measurements of Student Progress, an annual state assessment of math and reading given to students in Grades 3 through 8. This allows us to observe a prior-year test score in reading and math for each student in Grades 4 through 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. In elementary grades, we use the average of each student's score in math and reading, and in secondary grades we use the student's subject-specific score (math or reading; see Tables 1 through 6).

Table 2
Exposure Rates to Low-VAM Teachers in Fourth-Grade Classrooms by Student Disadvantage Indicator and Decomposition of Differences (in percentages)

	By FRL Status			By URM Status			By Quintile of Prior Performance (Math and Reading Average)		
Exposure to Low-VAM Teacher	FRL	Non-FRL	Diff	URM	Non-URM	Diff	Lowest	Non-Lowest	Diff
State level	11.99	8.14	3.85*	11.98	9.32	2.66*	11.20	9.58	1.62*
Decomposition of difference									
District level	11.00	9.04	1.96*	11.20	9.57	1.63*	11.13	9.60	1.53*
School level	0.59	–0.54	1.14*	0.41	–0.13	0.54*	0.13	–0.04	0.18
Classroom level	0.40	–0.36	0.76*	0.37	–0.12	0.49*	–0.06	0.02	–0.08
By quartile of district disadvantage (FRL, URM, or low performance)									
Lowest quartile (most advantaged)	14.55	5.75	8.80*	10.17	9.37	0.81	6.05	5.85	0.20
2nd quartile	8.34	10.09	–1.75*	9.42	5.39	4.03*	7.94	8.78	–0.84
3rd quartile	12.50	9.86	2.63*	16.25	13.40	2.85*	8.57	8.22	0.35
Highest quartile (most disadvantaged)	12.31	7.24	5.07*	11.22	7.85	3.37*	18.06	17.24	0.82

Note. VAM = value-added model; FRL = free or reduced-price lunch; URM = underrepresented minority; Diff = difference.
**p* < .05 (two-sided *t* test).

Table 3
Exposure Rates to Low-WEST-B Teachers in Fourth-Grade Classrooms by Student Disadvantage Indicator and Decomposition of Differences

	By FRL Status			By URM Status			By Quintile of Prior Performance (Math and Reading Average)		
Exposure to Low-WEST-B Teacher	FRL	Non-FRL	Diff	URM	Non-URM	Diff	Lowest	Non-Lowest	Diff
State level	11.27	8.59	2.68*	10.85	9.64	1.21*	10.65	9.78	0.87
Decomposition of difference									
District level	10.94	8.97	1.97*	11.15	9.50	1.65*	10.72	9.75	0.97*
School level	0.17	–0.19	0.37	–0.57	0.26	–0.83*	–0.28	0.140	–0.38
Classroom level	0.16	-0.18	0.34	0.27	-0.12	0.39	0.20	-0.08	0.28
By quartile of district disadvantage (FRL, URM, or low performance)									
Lowest quartile (most advantaged)	14.67	9.62	5.04*	5.85	9.94	–4.10*	7.11	7.47	–0.36
2nd quartile	10.03	9.43	0.61	12.52	9.99	2.53*	7.27	9.90	–2.63*
3rd quartile	15.13	10.89	4.24*	8.72	6.63	2.10–	11.02	11.84	–0.82
Highest quartile (most disadvantaged)	6.56	2.62	3.94*	12.81	11.47	1.34	14.15	10.52	3.63*

Note. WEST-B = Washington Educator Skills Test–Basic; FRL = free or reduced-price lunch; URM = underrepresented minority; Diff = difference.
**p* < .05 (two-sided *t* test).

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 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.[5] Like many researchers (Anzia & Moe, 2014; Clotfelter et al., 2005; Kalogrides & Loeb, 2013;

Table 4
Exposure Rates to Low-Qualified Teachers by Grade Level and Student Disadvantage Indicator and Decompositions of Differences (in percentages)

	FRL Status						URM Status						Low Prior Performance					
	By Student FRL			Decomposition of Difference			By Student URM			Decomposition of Difference			By Quintile of Prior Performance			Decomposition of Difference		
Grade	FRL	Non-FRL	Diff	District	School	Class	URM	Non-URM	Diff	District	School	Class	Lowest	Non-Lowest	Diff	District	School	Class
4th grade																		
Novice (<2 years experience)	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	11.99	8.14	3.85*	1.96*	1.14*	0.76*	11.98	9.32	2.66*	1.63*	0.54*	0.49*	11.20	9.58	1.62*	1.53*	0.18	–0.08
Lowest decile WEST-B	11.27	8.59	2.68*	1.97*	0.37	0.34	10.85	9.64	1.21*	1.65*	–0.83*	0.39	10.65	9.78	0.87	0.97*	–0.38	0.28
7th-grade math																		
Novice (<2 years experience)	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 experience)	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	2.41*	–0.12	0.86*	1.68*
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	15.52	10.85	4.67*	3.86*	0.35	0.47	16.09	11.95	4.13*	3.10*	–0.12	1.15*	17.33	11.54	5.79*	3.59*	0.49	1.71*
9th-grade algebra																		
Novice (<2 years experience)	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 experience)	8.86	7.38	1.48*	0.95*	0.50*	0.03	8.94	7.69	1.25*	0.92*	0.44*	–0.12	8.81	7.71	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	12.02	9.72	2.30*	1.54*	0.09	0.67*	12.62	10.08	2.54*	2.07*	–0.05	0.52	11.57	10.43	1.14	2.22*	–1.32*	0.24

Note. VAM = value-added model; WEST-B = Washington Educator Skills Test–Basic; FRL = free or reduced-price lunch; URM = underrepresented minority; Diff = difference.
*$p < .05$ (two-sided *t* test).

Table 5
Average Qualifications of Teachers by Grade Level and Student Disadvantage Indicator and Decompositions of Differences

	FRL Status						URM Status						Low Prior Performance					
	By Student FRL			Decomposition of Difference			By Student URM			Decomposition of Difference			By Quintile of Prior Performance			Decomposition of Difference		
Grade	FRL	Non-FRL	Diff	District	School	Class	URM	Non-URM	Diff	District	School	Class	Lowest	Non-Lowest	Diff	District	School	Class
4th grade																		
Average experience	13.43	14.18	–0.74*	–0.17*	–0.45*	–0.12*	12.61	14.23	–1.62*	–1.00*	–0.38*	–0.25*	13.33	13.97	–0.64*	–0.16*	–0.17*	–0.31*
Average prior VAM	0.000	0.036	–0.035*	–0.020*	–0.011*	–0.005*	0.001	0.024	–0.023*	–0.015*	–0.005*	–0.002*	0.004	0.024	0.020*	–0.013*	–0.006*	–0.002
Average WEST-B	272.78	276.18	–3.40*	–2.46*	–0.75*	–0.19	272.89	275.03	–2.14*	–1.745*	–0.07	–0.332*	273.26	274.78	–1.52*	–1.37*	0.07	–0.22
7th-grade math																		
Average experience	12.03	13.49	–1.45*	–0.61*	–0.69*	–0.16*	11.48	13.23	–1.75*	–1.06*	–0.46*	–0.23*	11.97	13.10	–1.14*	–0.41*	–0.49*	–0.24*
Average prior VAM	–0.026	0.033	–0.059*	–0.025*	–0.011*	–0.024*	–0.029	0.018	–0.046*	–0.021*	–0.005*	–0.019*	–0.066	0.029	–0.095*	–0.023*	–0.013*	–0.060*
Average WEST-B	284.70	287.50	–2.80*	–2.04*	–0.31*	–0.45*	284.04	286.92	–2.89*	–2.29*	–0.27*	–0.325*	283.39	287.11	–3.73*	–1.18*	–1.17*	–1.38*
7th-grade reading																		
Average experience	12.69	13.44	–0.75*	–0.15*	–0.45*	–0.15*	12.09	13.41	–1.32*	–0.73*	–0.34*	–0.25*	12.82	13.20	–0.38*	–0.07*	–0.26*	–0.05
Average prior VAM	0.014	0.051	–0.037*	–0.019*	–0.004*	–0.013*	0.013	0.041	–0.028*	–0.014*	–0.003*	–0.011*	–0.018	0.051	–0.069*	–0.021*	–0.007*	–0.041*
Average WEST-B	271.68	276.06	–4.39*	–4.22*	–0.18	0.01	270.55	275.26	–4.70*	–4.06*	–0.42*	–0.23	271.10	274.97	–3.87*	–3.00*	–0.80*	–0.06
9th-grade algebra																		
Average experience	11.27	12.22	–0.95*	–0.64*	–0.22*	–0.09	10.85	12.07	–1.23	–1.26*	–0.03	0.06	11.54	11.83	–0.29*	–0.03	–0.10*	–0.16*
Average prior VAM	–0.84	0.008	–0.092*	–0.080*	–0.007*	–0.006*	–0.107	–0.010	–0.097*	–0.077*	–0.011*	–0.010*	–0.087	–0.018	–0.069*	–0.052*	–0.007*	–0.010*
Average WEST-B	289.08	288.81	0.27	–0.28*	0.53*	0.02	289.12	288.88	0.23	–0.07	0.42*	–0.11	288.88	288.97	–0.09	–0.21	0.01	0.11
10th-grade reading																		
Average experience	12.90	13.39	–0.50*	–0.40*	–0.14*	0.04	12.56	13.37	–0.81*	–0.79*	–0.10*	0.08	13.02	13.25	–0.23*	–0.13*	–0.02	–0.07
Average prior VAM	0.002	0.008	–0.006*	0.000	–0.001*	–0.004*	0.001	0.007	–0.005*	0.001	–0.002*	–0.004*	–0.004	0.008	–0.012*	–0.001	–0.001*	–0.010*
Average WEST-B	277.13	279.28	–2.14*	–1.45*	–0.21*	–0.49*	277.27	278.73	–1.46*	–0.98*	–0.01	–0.47*	276.48	278.91	–2.44*	–1.98*	0.46*	–0.92*

Note. VAM = value-added model; WEST-B = Washington Educator Skills Test–Basic; FRL = free or reduced-price lunch; URM = underrepresented minority; Diff = difference.
*$p < .05$ (two-sided t test).

Table 6
Exposure Rates to Highly Qualified Teachers by Grade Level and Student Disadvantage Indicator and Decompositions of Differences (in percentages)

	FRL Status						URM Status						Low Prior Performance					
	By Student FRL			Decomposition of Difference			By Student URM			Decomposition of Difference			By Quintile of Prior Performance			Decomposition of Difference		
Grade	FRL	Non-FRL	Diff	District	School	Class	URM	Non-URM	Diff	District	School	Class	Lowest	Non-Lowest	Diff	District	School	Class
4th grade																		
Experienced (>10 years)	57.28	60.08	−2.79*	0.10	−1.94*	−0.95*	53.49	60.50	−7.01*	−4.27*	−1.99*	−0.75*	56.55	59.43	−2.88*	−0.18	−1.14*	−1.56*
Highest decile prior VAM	9.43	10.57	−1.15*	−0.59*	−0.63*	0.08	9.91	10.06	−0.15	0.08	−0.19	0.04	9.64	10.15	−0.50	−0.37	−0.27	0.14
Highest decile WEST-B	10.21	14.34	−4.12*	−4.31*	0.12	0.06	9.99	13.12	−3.13*	−3.12*	−0.05	0.04	10.68	12.69	−2.01*	−2.81*	1.19*	−0.39
7th-grade math																		
Experienced (>10 years)	49.11	57.68	−8.57*	−3.37*	−4.52*	−0.68*	46.22	56.08	−9.86*	−6.16*	−2.78*	−0.92*	48.83	55.39	−6.57*	−2.57*	−3.04*	−0.95*
Highest decile prior VAM	6.61	12.35	−5.74*	−2.65*	−1.26*	−1.83*	5.84	11.02	−5.19	−3.17*	−0.59*	−1.43*	4.53	11.46	−6.93*	−2.27*	−1.22*	−1.43*
Highest decile WEST-B	15.96	23.47	−7.51*	−6.77*	0.18	−0.92*	16.00	21.17	−5.17	−4.51*	−0.15	−0.50	15.30	21.34	−6.04*	−4.77*	−0.27	−1.00*
7th-grade reading																		
Experienced (>10 years)	56.60	59.95	−3.34*	−1.09*	−1.50*	−0.75*	53.55	59.92	−6.37*	−3.71*	−1.19*	−1.47*	56.68	59.01	−2.33*	−0.61*	−0.71*	−1.00*
Highest decile prior VAM	8.01	11.77	−3.76*	−2.26*	−0.30*	−1.21*	7.77	10.83	−3.06*	−2.22*	−0.14	−0.70	6.55	11.25	−4.70*	−1.75*	−0.20	−2.76*
Highest decile WEST-B	10.18	13.65	−3.47*	−3.81*	0.62*	0.28	8.33	13.40	−5.07*	−5.53*	0.32	0.14	8.91	13.09	−4.18*	−3.51*	−1.28*	0.61
9th-grad algebra																		
Experienced (>10 years)	48.27	52.23	−3.97*	−2.79*	−1.07*	−0.10	46.27	51.73	−5.46*	−5.84*	0.16	0.22	48.44	50.94	−2.50*	−0.51	−0.36	−1.64*
Highest decile prior VAM	7.56	12.38	−4.82*	−3.98*	−0.41*	−0.43	6.81	11.29	−4.48*	−3.19*	−0.91*	−0.38	6.85	11.19	−4.34*	−3.22*	−1.03	−0.38
Highest decile WEST-B	10.72	9.44	1.28*	0.04	1.62*	−0.39	11.36	9.59	1.77*	0.71	0.94*	0.12	9.51	10.30	−0.79	−0.86*	0.73*	−0.66
10th-grade reading																		
Experienced (>10 years)	57.14	59.19	−2.05*	−1.62*	−0.92*	0.49	55.62	59.14	−3.51*	−4.13*	−0.45*	1.07*	57.94	58.51	−0.56	−0.64*	−0.10	0.18
Highest decile prior VAM	9.17	10.60	−1.42*	0.45*	−0.51*	−1.36*	9.52	10.23	−0.71*	1.26*	−0.68*	−1.29*	7.78	10.63	−2.85*	0.83*	−0.40*	−3.28*
Highest decile WEST-B	13.15	17.29	−4.15*	−1.86*	−1.63*	−0.66*	13.84	16.10	−2.27*	−0.56	−0.81	−0.90*	12.28	16.48	−4.20*	−2.21*	−0.89*	−1.10*

Note. VAM = value-added model; WEST-B = Washington Educator Skills Test–Basic; FRL = free or reduced-price lunch; URM = underrepresented minority; Diff = difference.
*$p < .05$ (two-sided t test).

Koski & Horng, 2007), we use these detailed data to create an indicator for "novice teachers," which we define as those with 2 or fewer years of experience.[6] In extensions to our main analysis, we also consider average experience and an indicator for "experienced" teachers, which we define as those with greater than 10 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.[7] 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. In an extension, we also consider average WEST-B scores and an indicator for teachers whose scores fall in the top 10%.

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 & Hansen, 2010; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). 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:[8]

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

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 FRL, 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 teacher-effect estimates using empirical Bayes (EB) methods.[9] 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) 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.[10]

Method

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).[11] 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 $E_D(X_{ijkl})$ is the total number of disadvantaged students who have a low-quality teacher (summed over students, teachers, schools, and districts), and the denominator is the total number of disadvantaged students. Thus $E_D(X_{ijkl})$ is simply the percentage of disadvantaged students who are assigned to a low-quality teacher. We can also calculate the equivalent exposure rate for nondisadvantaged 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 nondisadvantaged students:

$$\text{Gap}_{\text{overall}} \equiv E_D(X_{ijkl}) - E_{ND}(X_{ijkl}).$$

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, whereas 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 $\text{Gap}_{\text{overall}}$ into three terms:

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

The first term, $\text{Gap}_{\text{district}} \equiv [E_D(\overline{X}_l) - E_{ND}(\overline{X}_l)]$, 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 nondisadvantaged 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, $\text{Gap}_{\text{school}} \equiv \{[E_D(\overline{X}_{kl}) - E_{ND}(\overline{X}_{kl})] - [E_D(\overline{X}_l) - E_{ND}(\overline{X}_l)]\}$, is the "school effect," or the difference in average school-level exposure rates to low-quality teachers between disadvantaged and nondisadvantaged students, subtracting out the difference in average district-level exposure rates. $\text{Gap}_{\text{school}}$ can be rewritten as $\{[E_D(\overline{X}_{kl}) - E_D(\overline{X}_l)] - [E_{ND}(\overline{X}_{kl}) - E_{ND}(\overline{X}_l)]\}$, which demonstrates that the school effect is also the difference in school-level rates of low-quality teachers between disadvantaged and nondisadvantaged students relative to the percentage of low-quality 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 are nondisadvantaged students within the same district.

The last term, $\text{Gap}_{\text{class}} \equiv \{[E_D(X_{ijkl}) - E_{ND}(X_{ijkl})] - [E_D(\overline{X}_{kl}) - E_{ND}(\overline{X}_{kl})]\}$, is the "classroom effect," which simply subtracts the sum of the school and district effects from the overall teacher quality gap. $\text{Gap}_{\text{class}}$ can be rewritten as $\{[E_D(X_{ijkl}) - E_D(\overline{X}_{kl})] - [E_{ND}(X_{ijkl}) - E_{ND}(\overline{X}_{kl})$, which demonstrates that the classroom effect is also the difference in exposure rates to low-quality teachers between disadvantaged and nondisadvantaged students relative to the percentage 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 are nondisadvantaged students within the same school.

Throughout our analysis, we investigate teacher quality gaps within one grade level at a time (instead of averaging across elementary grades, middle school grades, etc.) because of variation across grades in the distribution of the student and teacher characteristics we investigate. For example, student FRL enrollments drop in higher grades, and teacher characteristics, like WEST-B scores and experience, vary across grades (e.g., there are more novice seventh-grade math teachers than novice eighth-grade math teachers). If we averaged across grades, the teacher quality gaps could (conceivably) be due to variation across grades rather than within grades (which is what we want to investigate, e.g., are fourth-grade FRL students more likely to have a novice teacher than fourth-grade non-FRL students?).

Results

We present our results in three steps. First, to clarify our methods and take a close look at the inequitable distribution of the measures of low teacher quality across students in one "representative" grade level, we focus solely on fourth-grade classrooms and investigate the distribution of low-quality teachers across indicators of student disadvantage. Second, we repeat this procedure across representative grades for all three school levels (elementary school, middle school, and high school) and two subject areas (math and reading). This is intended to provide a complete picture of the inequitable distribution of "low-quality" teachers in Washington State. Finally, we extend these results by investigating the average characteristics of teachers and the distribution of high-quality teachers across indicators of student disadvantage.

Distribution of Low-Quality Teachers in Fourth-Grade Classrooms

Tables 1 through 3 give an overview of the distribution of low-quality teachers across all three indicators of student disadvantage—FRL, URM, and low prior performance (lower quartile prior-year test scores)—for fourth-grade classrooms in Washington State. In each table, the first row of results gives the exposure rates for disadvantaged and nondisadvantaged students ($E_D[X_{ijkl}]$ and $E_{ND}[X_{ijkl}]$ from Method section, respectively) for each indicator of disadvantage as well as the teacher quality gap ($\text{Gap}_{\text{overall}}$ from Method section). In Table 1, we can see that for each indicator of disadvantage, but particularly for URM students, disadvantaged fourth-grade students are more likely to be assigned to a novice teacher than are nondisadvantaged fourth-grade students (and each teacher quality gap is statistically significant at the .05 level). In Table 2, we see that disadvantaged fourth-grade students, and particularly students eligible for FRL, are more likely to be assigned to a teacher with low prior-year value added than are advantaged students. Finally, in Table 3, we see that both FRL and URM fourth-grade students are more likely to have a teacher with a low WEST-B score than are advantaged students.

Some interesting patterns emerge when we decompose these teacher quality gaps into district, school, and classroom effects (shown in the upper portion of Tables 1 through 3). The effects themselves are in the "Diff" column in Tables 1 through 3, while the terms defined in the Method section and used to calculate these effects—$E_D(\overline{X}_l)$ and $E_{ND}(\overline{X}_l)$ for the district effect, $E_D(\overline{X}_{kl}) - E_D(\overline{X}_l)$ and $E_{ND}(\overline{X}_l)$ and $E_{ND}(\overline{X}_{kl}) - E_{ND}(\overline{X}_l)$ for the school effect, and $E_D(X_{ijkl}) - E_D(\overline{X}_{kl})$ and $E_{ND}(X_{ijkl}) - E_{ND}(\overline{X}_{kl})$ for the classroom effect—are in the other columns. Across every indicator of student disadvantage and indicator of low teacher quality, the school and district effects are larger than the classroom effects, meaning that most of the inequitable sorting of disadvantaged fourth graders to low-quality teacher occurs across districts and schools rather than within schools.

However, the relative magnitudes of the district and school effects vary depending on the definition of student disadvantage and low teacher quality. For example, the gap in exposure to low-VAM teachers for low-performing fourth graders (see Table 2) appears to be driven primarily by sorting across districts; that is, low-performing fourth graders are much more likely to attend a district with a high percentage of low-VAM teachers than are higher-performing fourth graders. On the other hand, the gap in exposure to novice teachers for FRL fourth graders (see Table 1)

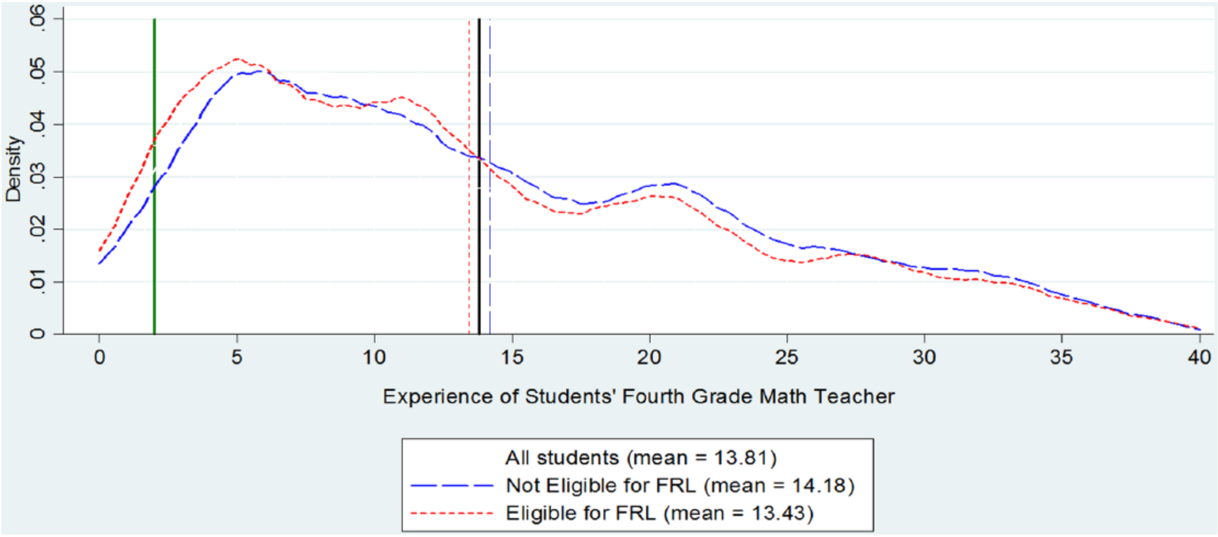

FIGURE 1. *Observed distribution of teacher experience in fourth-grade classrooms by student free/reduced-price lunch status*

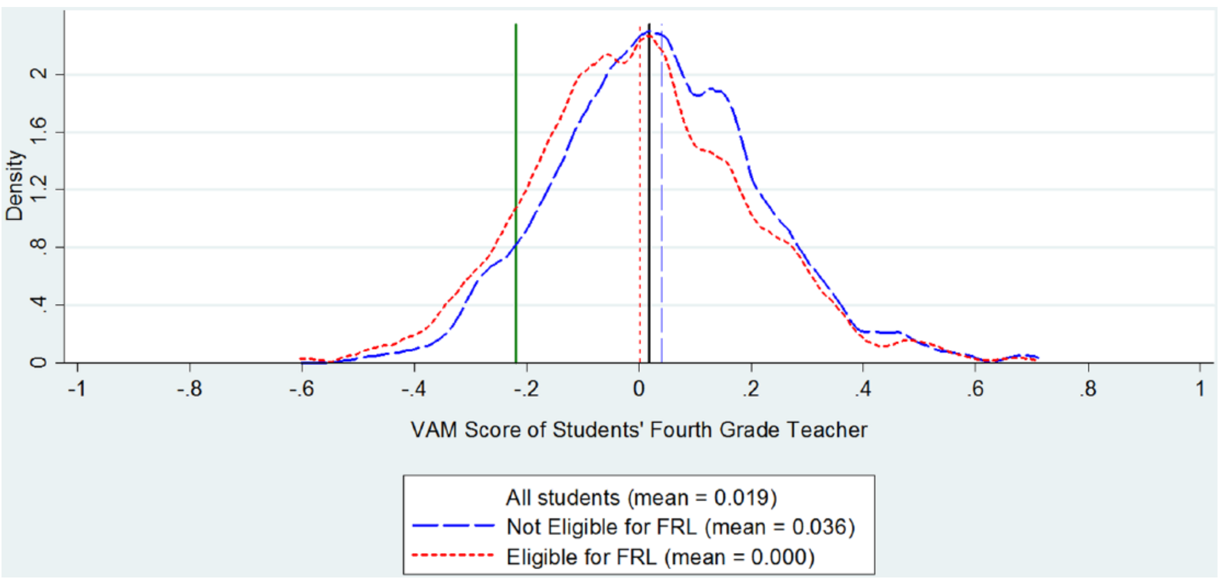

FIGURE 2. *Observed distribution of teacher value-added model score in fourth-grade classrooms by student free/reduced-price lunch status*

appears to be equally driven by sorting across districts and across schools within districts; that is, FRL fourth graders are more likely to attend a district with a high percentage of novice teachers than are non-FRL fourth graders in the state *and* are more likely to attend a school with a high percentage of novice teachers than are non-FRL fourth graders in the district. The only evidence for within-school sorting in fourth grade is for the exposure of URM students to low-VAM teachers (see Table 2); that is, URM fourth graders are more likely to be assigned to a low-VAM teacher than are non-URM fourth graders in the same school.

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 subsections. But before we proceed, we dig a little deeper into the inequitable distribution of low-quality teachers in fourth grade. First, Figures 1 through 3 show the observed distribution of teacher experience (Figure 1), teacher VAM (Figure 2), and teacher WEST-B scores (Figure 3) in fourth-grade classrooms by student FRL. In each figure, the first vertical line, to the left of all others, indicates our cutoff for "low-quality teachers," and the other vertical lines indicate the means for each group. In Figure 1, we see that the distribution of teacher experience for FRL fourth-grade students is weighted more heavily toward inexperienced teachers, and the average teacher experience for FRL fourth-grade students is almost a full year less than the average teacher experience for non-FRL fourth-grade students. Figure 2 demonstrates that the distribution of prior-year VAM estimates for teachers of FRL fourth graders is consistently lower than the distribution for non-FRL fourth graders, and the mean is 3.6% of a standard deviation lower. Finally, in Figure 3, we see that FRL fourth graders are both more likely to have a teacher with a low WEST-B score and less likely to have a teacher with a high WEST-B score. We will return to the distribution of "high-quality" teachers in an extension of our main results.

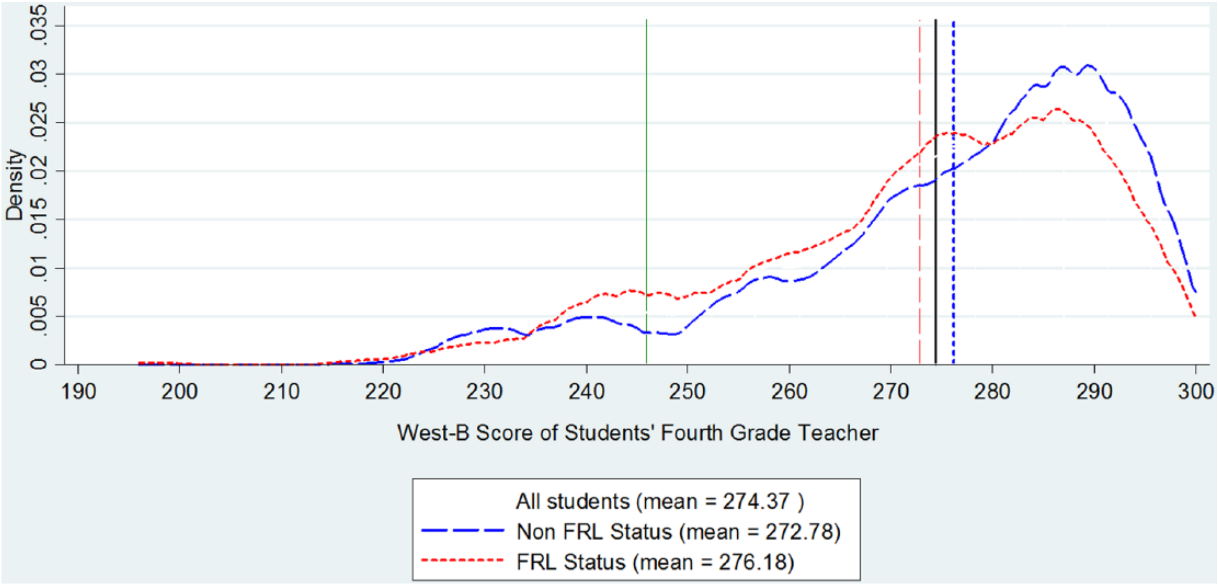

FIGURE 3. *Observed distribution of teacher Washington Educator Skills Test–Basic score in fourth-grade classrooms by student free/reduced-price lunch status*

Next, Figure 4 plots the exposure rate to novice teachers for fourth-grade FRL students against the exposure rate for fourth-grade FRL students within the 23 largest districts in the state.[12] Although the majority of districts fall above the 45-degree line—indicating the fourth-grade FRL students in these districts are more likely to be assigned a novice teacher than are fourth-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 and non-FRL students.

Finally, the lower portion of Tables 1 through 3 explores whether the teacher quality gap is higher in disadvantaged or more advantaged districts. Perhaps not surprisingly, we find that the distribution of low-quality teachers is most inequitable within the most disadvantaged districts. For example, in districts in the highest quartile of poverty in the state ("most disadvantaged"), FRL fourth graders are 2.40 percentage points more likely to have a novice teacher than are non-FRL fourth graders (see Table 1).

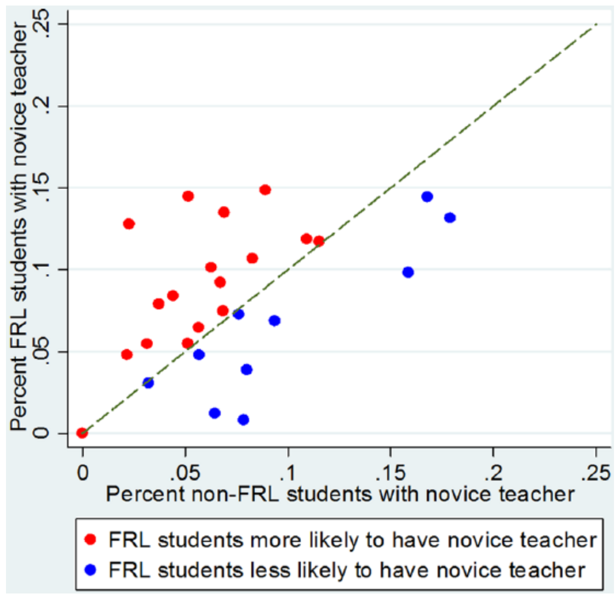

FIGURE 4. *Exposure rates to novice teachers in fourth-grade classrooms by student free/reduced-price lunch status for large districts*

Distribution of Low-Quality Teachers Across All Student Indicators and Grade Levels

We now proceed to investigate teacher quality gaps for every combination of school level, student disadvantage indicator, and indicator of teacher quality. Table 4 presents the overall teacher quality gap for low-quality teachers as well as the decompositions into district, school, and classroom effects.[13] The first three rows of results in Table 4 repeat the relevant results from Tables 1 through 3 about the distribution of novice, low-VAM, and low-WEST-B teachers across various indicators of student disadvantage in fourth grade. The remaining rows present the analogous results for other grade levels and subjects.

We first focus on the teacher quality gap for each of the combinations.[14] 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 are nondisadvantaged students in the same grade level. The only exceptions are the distribution of teachers with low credential exam scores across students by performance in fourth-grade classrooms, across all indicators of disadvantage in ninth-grade algebra classrooms, and by performance in 10th-grade reading 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 4. The highest gap is for the distribution of teachers with low prior VAM estimates across students in seventh-grade math with low prior performance; 19.25% of low-performing seventh-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 seventh 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 seventh-grade reading. But large teacher quality gaps exist throughout Table 4, reinforcing the magnitude of the inequitable distribution of teacher quality across student subgroups in Washington State.

That said, we believe that even relatively "small" teacher quality gaps in Table 4 are policy relevant because they represent large differences in relative probabilities. For example, FRL seventh graders in Washington are "only" 1.82 percentage points more likely to have a novice reading teacher than are non-FRL seventh graders (6.67% vs. 4.85%). But this means that FRL seventh graders in Washington are 37.5% more likely to have a novice reading teacher than are non-FRL seventh graders, which is a substantial increase in relative probability terms.

We next turn our attention to the decomposition of each of these teacher quality gaps. For nearly every combination of school level, student disadvantage indicator, and indicator of low teacher quality, the district effect is positive and statistically significant (i.e., disadvantaged students are more likely to attend districts with high percentages of low-quality teachers than are nondisadvantaged students in the same grade). In many cases (e.g., student FRL and teacher WEST-B in seventh-grade reading, student low performance and teacher VAM in ninth-grade algebra, and student URM and teacher WEST-B scores in 10th-grade reading), the district effect explains the majority of the teacher quality gap, meaning that—in these cases—student and teacher sorting across districts is primarily responsible for the inequitable distribution of teacher quality across students.

The school effect is also positive and statistically significant for most combinations of school level, student disadvantage indicator, and indicator of low teacher quality. In fact, the school effect is the largest effect explaining the inequitable distribution of novice teachers across FRL students in seventh-grade math and for FRL and URM students in seventh-grade reading. In other words, student and teacher sorting across schools within a district appears to be a major factor in explaining inequities at the middle school level. One explanation for this is that high-poverty and high-minority middle schools have a more difficult time attracting teachers than more advantaged schools in the same district, so they must hire a disproportionate number of novice teachers to staff their classrooms.

Finally, the classroom effect is positive and statistically significant for about half of the combinations of school level, student disadvantage indicator, and indicator of low teacher quality. The two largest classroom effects occur at the middle school level: Seventh graders with low prior performance are 7.49 percentage points more likely to have a low-VAM teacher in math and 6.97 percentage points more likely to have a low-VAM teacher in reading than are other students in the same school. This suggests that low-performing seventh graders may be disproportionately "tracked" into classrooms with previously ineffective teachers.

Average Distributions and Distribution of High-Quality Teachers

To this point, we have focused exclusively on the distribution of low-quality teachers, with the justification that having a low-quality teacher can influence a student's academic trajectory. However, research has also shown that differences throughout the distribution of teacher quality can have long-term impacts on students (Chetty et al., 2013). Therefore, we also examine differences in average teacher quality across indicators of student disadvantage in Table 5 and the distribution of high-quality teachers—defined as teachers with more than 10 years of experience or in the top decile of the VAM or WEST-B distributions—in Table 6.

In Table 5, we see that patterns in the distribution of low-quality teachers largely hold for average teacher characteristics. That is, across nearly every combination of school level, student disadvantage indicator, and indicator of low teacher quality, disadvantaged students have lower-quality teachers compared to advantaged students, on average. The magnitudes of some of the average teacher quality gaps are striking; for example, the average experience of a URM seventh grader's math teacher is 1.75 years less than the average experience of a non-URM seventh grader's math teacher, and the average WEST-B score of a FRL seventh grader's reading teacher is 4.39 points less than the average WEST-B score of a non-FRL seventh grader's reading teacher. The patterns in the decomposition of the teacher quality gaps, discussed for low-quality teachers in Table 4, also generally hold for average teacher characteristics.

Finally, in Table 6, we see that disadvantaged students are considerably less likely to be taught by a high-quality teacher in nearly every grade level. In fact, the magnitudes of the teacher quality gaps for exposure to high-quality teachers are generally larger than the teacher quality gaps for exposure to low-quality teachers; that is, disadvantaged students may be more likely to be taught by a low-quality teacher, but they are even less likely to be taught by a high-quality teacher. That said, two results in Table 6 run counter to this trend: In ninth-grade algebra classrooms, FRL and URM students are more likely to be taught by a teacher in the top decile of the WEST-B distribution than are non-FRL or non-URM students. In each case, sorting of students and teachers across schools explains the majority of this difference. One potential explanation for this, then, is that districts are finding a way to get their high-quality algebra teachers into disadvantaged schools.

Discussion and Conclusions

Our findings provide comprehensive, descriptive evidence 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—FRL, URM, and low prior academic performance—at virtually every school level in Washington State. This descriptive analysis is important because it both quantifies the magnitudes of the teacher quality gaps in Washington's public schools and highlights the grades and subjects—such as seventh-grade math and reading—where these gaps are particularly pronounced.

We also demonstrate that, regardless of whether we consider indicators of low teacher quality, average teacher quality, or high teacher quality, 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, Gross, & Player, 2010; Guarino, Santibañez, & Daley, 2006; Hanushek, Kain, & Rivkin, 2004; Scafidi, Sjoquist, & Stinebricker, 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, Lankford, Loeb, & Wyckoff, 2013; Engel, Jacob, & Curran, in press), 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.[15] And new evidence (Goldhaber, Krieg, & Theobald, in press) 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, because 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 article (Goldhaber, Lavery, & Theobald, 2014), 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.

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, Glennie, Ladd, and Vigdor (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%. Grissom, Loeb, and Nakashima (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.

As policymakers encounter mounting evidence that the most disadvantaged students are much more likely than their peers to face low-quality teachers, they will need to define their "ideal" distribution of teacher quality. Some might argue that, in an ideal world, teacher quality would be equitably distributed across all students; that is, a disadvantaged student would have the same probability of receiving a low-, average-, or high-quality teacher as an advantaged student. Others might argue that, to combat the well-documented achievement gaps that exist in public schools, disadvantaged students should actually have greater access to higher-quality teachers than should advantaged students. Although we are reluctant to insert our own ideals, we doubt that anyone would argue that the current distribution of teacher quality—and specifically, the meaningful and policy-relevant teacher quality gaps between advantaged and disadvantaged students that we document throughout this article—is in the best interests of all students.

NOTES

This research was made possible in part by generous support from the Bill and Melinda Gates Foundation, CALDER, and an anonymous foundation and has benefited from helpful comments from participants in the Causal Inference Working Group at the University of Michigan, the Center for Education Policy Analysis Seminar at Stanford University, and the Association for Education Finance and Policy 2014 Conference. We also thank Christopher Tien and Malcolm Wolff for excellent research assistance. The statements made and views expressed in this article do not necessarily reflect those of the study's sponsors or the institutions with which the authors are affiliated. Any and all errors are solely the responsibility of the authors.

[1]A follow-up article by the same authors (Clotfelter, Ladd, Vigdor, & Wheeler, 2006) explores the sorting of teachers and principals to high- and low-poverty schools.

[2]See Goldhaber and Theobald (2012) for a full discussion of this issue.

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

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

[5]The S-275 contains the experience that teachers are credited with for pay purposes, which may not include out-of-state teaching, teaching in a private school, or substitute teaching.

[6]We choose to focus on the distribution of early-career teachers because it is well known that teachers become more effective over the first few years of their career (e.g., Rice, 2013).

[7]Teachers may take the test multiple times to get a passing score on all three tests, so we use the test scores from the first time each teacher took the Washington Educator Skills Test–Basic. This ensures that

teachers taking the test for the fifth time, for example, are not judged as "comparable" to teachers who passed all three tests on the first attempt.

[8]The base model is very similar to the model estimated by Isenberg et al. (2013). We make slight modifications to this model to estimate value-added models in high school. For math, we estimate only the model for ninth-grade students enrolled in algebra who took the algebra End-of-Course exam at the end of the year (using eighth-grade scores as prior-year test scores). For reading, the dependent variable is the student's High School Proficiency Exam score in 10th grade. However, students are not tested in reading in ninth grade, so the prior-year test scores are the student's eighth-grade test scores. We then include two teacher fixed effects—one for the ninth-grade reading teacher and one for the 10th-grade reading teacher—to account for combined contributions to the student's 10th-grade test score.

[9]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., 2004). But recent evidence (Herrmann, Walsh, Isenberg, & Resch, 2013) also suggests that shrinkage improves the estimates for teachers' "hard-to-predict" students. We use the standard approach that has been commonly estimated in the literature (an appendix on empirical Bayes shrinkage is available from the authors by request).

[10]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 & Loeb, 2013).

[11]We use indicators of low teacher quality in our main results but also consider indicators of high quality and continuous measures of quality in extensions.

[12]We use an arbitrary cutoff of districts with at least 500 fourth-grade teachers to ensure adequate sample sizes for this figure.

[13]We present results for fourth-grade in elementary school, seventh-grade math and reading in middle school, and ninth-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.

[14]We also examine teacher quality gaps for more precise race/ethnicity indicators and for English language learners (ELLs). We find consistent teacher quality gaps for Hispanic, Black, American Indian, and ELL students relative to their more advantaged peers; full results are available upon request. Results for novice teachers are also robust to the cutoff used for novice teacher (i.e., when we consider only 1st-year teachers or teachers with 5 or fewer years of experience rather than the definition (less than 2 years) used throughout the article.

[15]See Goldhaber, Krieg, Theobald, and Brown (in press) for more discussion of the consequences of Washington State's single salary schedule.

REFERENCES

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*, 95–135.

Anzia, S. F., & Moe, T. M. (2014). Collective bargaining, transfer rights, and disadvantaged schools. *Educational Evaluation and Policy Analysis*, *36*, 83–111.

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*, 83–117.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013).*Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood* (No. w19424). Cambridge, MA: National Bureau of Economic Research.

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*, 1352–1370.

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*, 377–392.

Clotfelter, C., Ladd, H. F., Vigdor, J., & Wheeler, J. (2006). High-poverty schools and the distribution of teachers and principals. *North Carolina Law Review*, *85*, 1345–1379.

Engel, M., Jacob, B., & Curran, F. (in press). 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 H. Ladd & E. Fiske (Ed.), *Handbook of research in education finance and policy* (pp.146–165). New York, NY: Routledge.

Goldhaber, D., & Brewer, D.J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Education Evaluation and Policy Analysis*, *22*, 129–145.

Goldhaber, D., Gross, B., & 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*, 57–87.

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

Goldhaber, D., Krieg, J., & Theobald, R. (in press). Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. *Economics of Education Review*.

Goldhaber, D., Krieg, J., Theobald, R., & Brown, N. (in press). The line is shorter over here: The probability and time to first job of prospective teachers with different training. *Phi Delta Kappan*.

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

Goldhaber, D., & Theobald, R. (2012). *Do different value-added models tell us the same things?* Stanford, CA: Carnegie Knowledge Network.

Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy*, *1*, 28–39.

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*, 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*, 173–208.

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

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

Herrmann, M., Walsh, E., Isenberg, E., & 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*, 690–717.

Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). *Access to effective teaching for disadvantaged students*. Washington, DC: 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*, 304–316.

Kalogrides, D., Loeb, S., & Beteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, *86*, 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*, *2*, 262–300.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, *33*, 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*, 37–62.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*, 67–101.

Player, D. (2010). Nonmonetary compensation in the public teacher labor market. *Education Finance and Policy*, *5*, 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*, 332–348.

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

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*, 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*, 104–122.

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

U.S. Department of Education. (2010). *Public elementary/secondary school universe survey, 2000–01, 2005–06, 2007–08, and 2008–09*. Washington, DC: National Center for Education Statistics, Common Core of Data.

AUTHORS

DAN GOLDHABER, PhD, is the director of CALDER at the American Institutes for Research, 3876 Bridge Way N., Seattle, WA 98103; *dgoldhab@uw.edu*. His research focuses on teacher quality and teacher labor markets.

LESLEY LAVERY, PhD, is an assistant professor at Macalester College, 1600 Grand Ave., St. Paul, MN 55105; *llavery@macalester.edu*. Her research focuses on collective bargaining, school choice, and policy implementation.

RODDY THEOBALD is a researcher at CALDER at the American Institutes for Research, 3876 Bridge Way N., Suite 201, Seattle, WA 98103; *roddy.theobald@gmail.com*. His research focuses on the teacher labor market and its implications for student outcomes.

Manuscript received March 25, 2014
Revisions received June 26, 2014, and May 22, 2015
Accepted June 1, 2015