HCEO WORKING PAPER SERIES

Match Effects and the Gains from Alternative Job Assignments: Evidence from a Teacher Labor Market

Mariana Laverde
Elton Mykerezi
Aaron Sojourner
Aradhya Sood

Working Paper

2025-001

01-2025; 02-2026

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Match Effects and the Gains from Alternative Job Assignments: Evidence from a Teacher Labor Market *

Mariana Laverde — Boston College

Elton Mykerezi — University of Minnesota

Aaron Sojourner — W. E. Upjohn Institute

Aradhya Sood — University of Toronto

February 3, 2026

Abstract

This paper studies the relative importance of teacher-student match effects and general teacher effectiveness in producing student learning, and quantifies gains from alternative teacher assignments. We estimate a framework that separates these components, allowing match quality to vary with observable student characteristics and unobservable teacher-school factors. Using more than a decade of administrative data from a large urban district, we address endogenous sorting with quasi-random assignment variation induced by differences in driving time between teachers and schools. Match effects are similar in magnitude to general effectiveness. Teacher-acceptable reassignments can raise average test scores by about 0.13 standard deviations.

Keywords: Teacher effectiveness; Teacher–student match effects; Assignment and sorting; Education production; Labor markets in education

JEL Codes: I21, J45, I24, J24.

*This paper circulated previously under the title "Gains from Alternative Assignment? Evidence from a Two-Sided Teacher Market". This paper benefited from discussions with Nikhil Agarwal, Barbara Biasi, Jesse Bruhn, Chris Campos, Lucas Coffman, William Delgado, Michael Dinerstein, Chao Fu, Michael Grubb, Peter Hinrichs, Caroline Hoxby, Adam Kapor, Charlie Murry, Minseon Park, Theo Papageorgiou, Jonah Rockoff, Stephen L. Ross, Douglas Staiger, Sukjoon Son, Yang Song, Rich Sweeney, Chris Walters, Roman A. Zarate, Seth Zimmerman and seminar participants at the Barcelona School of Economics Summer Forum Workshop on Applied Industrial Organization, International Industrial Organization Conference, NBER Education Meetings, Northeast Economics of Education Workshop at Brown University, Stanford Institute for Theoretical Economics (SITE) Empirical Market Design Conference, Southern Economic Association Conference, Society of Labor Economists Conference, Texas A&M IO Day, Toulouse School of Economics - Economics of Education Workshop, Universidad de Montevideo - School Choice Workshop, Washington University at St. Louis - Empirical Analysis of School Choice Systems Workshop, and Western Economic Association Conference. This project relied on financial support from the McKnight Foundation. This study falls under the University of Minnesota IRB's project 1510S79046. Emails: laverdem@bc.com, myker001@umn.edu, sojourner@upjohn.org, and aradhya.sood@rotman.utoronto.ca.

1. Introduction

Each year in the United States, roughly 50 million public-school students are assigned to about 3.8 million teachers (Irwin et al., 2024). These teacher-student matches emerge from a sequence of largely decentralized and often uncoordinated decisions. Teachers choose which positions to apply for and accept, schools decide whom to hire, and families select which school to attend. This assignment process is far from innocuous on students' learning outcomes. Teachers are one of the most influential school-based factors in student learning (Hanushek, 2020; Kane and Staiger, 2008; Hanushek and Rivkin, 2012; Jackson, 2018), and exposure to a highly effective teacher can generate substantial gains in students' long-run outcomes, including adult earnings (Chetty et al., 2014b). This raises the question of how much optimizing the assignment of teachers to classrooms could improve student outcomes. Using data from a large urban school district in the U.S. Midwest, this paper examines what student learning gains can be realized by matching teachers to the positions where they are most effective–or, equivalently, to the jobs with the highest match effects.

The potential gains from an alternative assignment of teachers to classrooms depend critically on the distribution of match effects–the horizontal component of teacher effectiveness–and its dispersion relative to that of teacher general effectiveness–the vertical component. If variation in general effectiveness dominates, alternative assignments offer limited aggregate benefits. Conversely, if match effects are substantial, optimizing teacher assignments could generate significant gains in average student learning. Accurately measuring match effects is therefore essential. Policy proposals such as "deselection" of low value-added teachers or moving high value-added teachers to disadvantaged schools assume a minimal role of match effects (Hanushek et al., 2009; Staiger and Rockoff, 2010; Grissom et al., 2014; Chetty et al., 2014a). When match effects are present, however, reassignment or dismissal policies that ignore the horizontal component of teacher effectiveness may fail to deliver expected gains (Jackson, 2013; Condie et al., 2014).

In this paper, we study the relative importance of match effects and general teacher effectiveness in the production of student learning. These two components are typically confounded in observed value-added measures. To avoid this, we exploit quasi-random variation in teacher assignments to separately identify each. We use the resulting estimates to quantify the learning gains attainable from reassigning teachers based on their match effects, with particular attention to alternative assignments that would be acceptable to teachers under existing compensation policies.

Prior work has typically studied teacher-student match effects along a single, often dichotomous, student characteristic–such as race, income, or baseline achievement–considered in isolation (Dee, 2005; Biasi et al., 2022; Delgado, 2023; Graham et al., 2023; Bates et al., 2024).[1] This literature has established that teachers can exhibit meaningful match effects along several dimensions. If multiple characteristics matter, however, a natural question arises as to whether additional–and potentially interacting–student attributes also shape match effects. Expanding the set of student types by incorporating cross-products of characteristics quickly becomes infeasible in finite samples, as cell sizes shrink rapidly. Moreover, if match effects operate along dimensions that are unobservable to the researcher, even a richly parameterized model would fail to capture them. To address these limitations, this paper develops a framework that flexibly accommodates match effects arising from both observable student characteristics and unobservables varying at the teacher–school level, allowing us to measure the distribution of match effects while summarizing their multidimensional nature.

Because the model places particular emphasis on unobservable match effects, we explicitly address the possibility that teachers select into schools in ways correlated with these effects. We extend the model to jointly capture teachers' decisions about where to work and student outcomes, allowing for correlation between the two (Roy, 1951). Empirically, we leverage more than a decade of data from a large urban school district in the United States, including detailed student-teacher matches, student test scores, and information from a centralized teacher reassignment market. Identification exploits quasi-random variation in teacher assignments induced by a shifter to teachers' job location choices—differences in driving time between teachers and schools—which generates plausibly exogenous movement across assignments, a claim we assess empirically in the

[1] Ahn et al. (2024) and Umosen (2024) are exceptions by studying multiple characteristics simultaneously.

data. This variation allows us to recover both the dispersion of unobservable match effects and their correlation with teacher sorting behavior.

We find that match effects exhibit substantial variation, comparable in magnitude to that of general effectiveness, implying meaningful scope for average gains from alternative teacher assignment. Focusing on math learning outcomes, our estimates indicate that, when evaluated over all feasible teacher-student assignments, a one-standard deviation (SD) increase in teacher general effectiveness raises average student test scores by 0.12 SD. By comparison, a one-standard deviation increase in match effects associated with observable student characteristics increases test scores by 0.08 SD, while a similarly sized increase in match effects unrelated to observables raises scores by 0.07 SD for an average student. Together, these estimates highlight the importance of accounting for match effects that operate beyond observable student characteristics.

Our estimates also reveal negative Roy selection in teacher assignments, operating primarily through observable student characteristics. In particular, teachers who are estimated to be more effective with students who are not proficient tend to sort toward classrooms with more proficient students, generating systematic mismatch along the proficiency margin. By contrast, we find positive selection along race and ethnicity: teachers who are most effective teaching Black or Hispanic students are more likely to sort toward classrooms serving those students. Because our estimates indicate that the scope for match effects is larger along the proficiency dimension, these opposing patterns imply an overall negative Roy selection in teacher-position assignments.

Under a policy that reallocates teachers across schools and classrooms, imposing no layoffs and requiring that no teacher assigned under the counterfactual be made worse off relative to her observed assignment, a fully informed planner could raise average test scores by about 0.13 standard deviations, holding student classroom composition fixed. For the average student, this gain corresponds to an increase in lifetime earnings with a present value of $7,300 at age 12 (Chetty et al., 2014b).[2] While this alternative assignment

[2]Our sample is closer in age to 7 years old. The corresponding lifetime earnings gain at this age is approximately $6,300. These calculations assume that percentage gains in earnings remain constant over the life cycle and are discounted at a 3 percent real rate to age 12 or 7, respectively.

benefit both high– and low–achieving students, as well as White and BIPOC students, gains are largest for higher–achieving and White students, thereby widening achievement gaps. In contrast, a policy that instead maximizes the share of proficient students yields similar, though somewhat smaller, average test score gains while mitigating increases in racial and achievement inequality.

A decomposition of average test score gains shows that, under a policy that maximizes average achievement, unobservable match effects drive about 70 percent of the aggregate gains. Assigning more generally effective teachers to larger classrooms accounts for roughly 26 percent of the gains, with the remaining share attributable to matching on observable characteristics. The dominance of unobservable match effects reflects the fact that gains from matching on observables are attenuated by classroom composition. Because classrooms enroll students of heterogeneous types, the potential gains associated with observable match effects cannot be fully realized. As a result, unobservable dimensions of fit or match effects account for most of the unrealized gains from improving teacher–student assignments, underscoring the importance of investing in policy tools that help teachers, principals, and school districts learn where individual teachers are most productive.

Related Literature: Our paper relates to a growing literature documenting that teachers exhibit heterogeneous effectiveness across students, implying scope for gains from improved assignment (Condie et al., 2014; Aucejo et al., 2022; Biasi et al., 2022; Delgado, 2023; Bates et al., 2024; Bobba et al., 2024; Ahn et al., 2024; Umosen, 2024; Eastmond et al., 2025). Much of this work models match effects along a single salient student dimension—such as prior achievement or disadvantage—or, when allowing richer heterogeneity, restricts attention to observable student characteristics only (e.g., Ahn et al. (2024); Umosen (2024)), abstracting from unobservable match effects that may systematically affect teacher effectiveness. In contrast, our framework estimates match effects by allowing teacher effectiveness to vary flexibly with a rich set of observed student characteristics as well as with unobserved features of the teacher–school match.[3]

Relative to the existing teacher labor market literature, our paper emphasizes the

[3]Related work addressing unequal access to teachers without explicitly modeling teacher–student match effects includes Combe et al. (2022).

potential correlation between teachers' assignment decisions and student outcomes. This idea was previously studied by Jackson (2013), who examines whether teachers sort toward positions in which they have a comparative advantage, using a reduced–form approach that does not model teacher utilities or explicit choice behavior. While Jackson (2013) finds evidence of positive sorting on average, our analysis—leveraging data on teachers' application and acceptance decisions jointly with student outcomes—reveals a more nuanced pattern: teachers sort in ways that are both positively and negatively correlated with their match effects. In particular, we document positive selection on race and negative selection on student proficiency, implying that, in our context, these forces combine to generate an overall pattern of negative sorting.

The paper also contributes to the literature studying policy interventions aimed at improving student achievement. One strand of this literature focuses on "deselection" policies that propose removing teachers at the bottom of the effectiveness distribution, where effectiveness is typically measured using homogeneous value–added models (Staiger and Rockoff, 2010; Rothstein, 2015). By accounting for teacher–student match effects, beyond average effectiveness, and by considering alternative assignment policies in addition to terminations, our counterfactuals encompass a broader and more nuanced set of policy options. A second strand examines the effects of reallocating students across schools within a district, often finding limited gains once capacity constraints and vertical school quality differences are taken into account (e.g., Abdulkadiroğlu et al., 2025). We complement this work by emphasizing that a substantial share of estimated heterogeneity in teacher effectiveness occurs within rather than between schools (Chetty et al., 2014a), implying that reallocating students while holding teachers fixed may understate the scope for match gains. By instead reallocating teachers across positions while holding student composition fixed, our analysis targets a margin along which heterogeneity may be empirically larger, and match effects may therefore have more pronounced impacts on student achievement. This approach may be particularly relevant given the costs of moving students for families and districts (Laverde, 2024; Angrist et al., 2024; Idoux, 2022; Campos et al., 2025), and the mixed evidence on the achievement impacts of student reassignment (Abdulkadiroğlu et al., 2025; Campos and Kearns, 2024; Angrist et al., 2024; Deming, 2011;

Deming et al., 2014).

2. Setting, Data, and Empirical Evidence

2.1 Setting

Our study focuses on a large, diverse urban school district in the U.S. Midwest. According to our analysis of data from the National Center on Teacher Quality (2022), seven in ten large U.S. school districts, including the district we study, delegate authority to school leadership to choose who will fill open teaching positions at the school. Open positions can arise because of the creation of new positions, retirements, or quits or firings of incumbent teachers. Incumbent teachers within the school district typically fill most open positions through the district's internal transfer mechanism. External candidates can also fill these vacancies, but only after incumbent candidates are considered, as specified in the collective bargaining agreement of the school district.[4]

The school district that we study created a centralized ITS to govern its matching process and to serve as a clearinghouse for its internal labor market. The process comprises two successive rounds of applications, interviews, and offers in the spring of each academic year. Based on the district projections of school enrollments, budgets, and incumbent teachers' commitments to retire or take leave, each school projects and posts its vacancies on the ITS for the coming year. Any incumbent teacher can apply to any vacancies within the school district for which her licenses qualify her.

After the application window closes, the school district checks each applicant's eligibility for the positions to which she applied, and the system automatically grants interviews to the four most senior applicants for each vacancy, per the collective bargaining agreement. Then, for each vacancy, each school principal reviews these automatic interviewees and the remaining applicant pool and can choose up to four additional applicants for interview, yielding a maximum of eight interviewees per position.[5] When principals eval-

[4]At least 40% of large U.S. school districts explicitly prioritize internal transfer candidates over external hires when filling open positions (National Center on Teacher Quality, 2022).

[5]While schools can abstain from interviewing any of the four most senior applicants, they would lose the option to make an offer to any applicant outside this group. If they decide to interview only a subset of the most senior applicants, they have to invite them in order of seniority.

uate an applicant for the interview process, they observe the applicant's CV. The district recommends a standard CV format that includes details about the applicant's education, employment history, and other qualifications.

By the end of the interview period, each principal can submit a ranking of up to four interviewees to the district via the ITS. Next, the system automatically and simultaneously emails offers to the first-ranked interviewee for each position in the district. Thus, an applicant can receive zero, one, or multiple offers at this stage. Offerees have 48 hours to accept up to one offer. After 48 hours, the ITS automatically withdraws unaccepted offers and emails an offer to the second-ranked interviewee for each open position. This process repeats until each position's ranked list is exhausted or all positions are filled. Within a round, no teacher can renege on a previously accepted offer. After the first round is completed, any vacancies that remain or new vacancies that arise from transfers during the first round can be posted in the second round. The whole application, interview, ranking, and offer process repeats again. After the second round, remaining vacancies become open to both internal and external candidates and are filled in a decentralized manner. We focus on data from the first two rounds, for which detailed information on all participants' actions is observed.

2.2 Administrative Data

We use data from the ITS spanning 2010 to 2019 and merge these records with student outcome data for the same years.[6] In particular, we observe the vacancy postings, applications, interview decisions, rankings, offers, and acceptances. In addition, for each teacher, we observe seniority rank, experience, education, ethnicity, race, gender, and current position assignment for every year. We also observe each teacher's home address each year, allowing us to measure the driving time to each open position in a teacher's choice set.[7] Because we observe teacher licenses, we can accurately measure each teacher's choice set each year. Accurately defining the choice set is important, as inaccurately specifying choice sets leads to biased utility estimates (Almagro and Sood, 2025). Our ability to

[6] We also use student outcome data from 2009 to measure prior-year achievement in the student outcomes model.

[7] We geocode teacher and school addresses and measure driving times using the Google Maps API.

measure this is uncommon in the literature.

We include the positions that the district evaluates for math and reading effectiveness and exclude others[8]. We restrict attention to grades 4–8 to measure teacher effectiveness most reliably. This restriction is standard in the literature because (1) job assignment in these grades provides a strong match to tested students and (2) students are mandated to take standardized tests in these grades and the prior grades. We drop teachers working less than half time to ensure that a teacher spends a significant number of hours with the students in their classes.

The baseline sample for analysis contains 823 teachers. On average, we observe a teacher for an average of 4 years, with a total of 3,268 teacher–year observations in our sample (Table 1: Panel A). Just over half the teachers ever participated in the centralized ITS by applying to at least one open position between 2010 and 2019. Three-fourths of the teachers are women. About 4 in 5 are White, followed by 9 percent who are Black. Teachers average 13.4 years of experience and 5 years of higher education. Teachers who apply for transfers are less experienced, on average. Teachers who seek a transfer apply to an average of 6.7 positions per year-round out of 36.5 positions for which they are eligible (Panel B). On average, applicants receive interviews for 3.4 positions, are ranked by principals for 1.5 positions, and receive 0.8 offers.

On the other side of the market, the ITS included 972 vacancy postings for grades 4–8 from 2010 to 2019.[9] Though 213.7 teachers are eligible for the average posting, each position, on average, attracted approximately 5 applicants (Panel C). The propensity of qualified teachers not to apply to open positions motivates the inclusion of an inertia term in our model. An average of 2.5 applicants were interviewed and 0.6 offers made for each position.

For each student in each academic year, we observe current and prior year state-

[8]To identify which positions to include in our analysis, we rely on the district's internal value-added evaluation system. We classify a position as evaluated in math and/or reading if the teacher assigned to that position in that year receives a reading and/or math value-added score. We apply this rule starting in 2013, when the district adopted the value-added system. For earlier years, we use the assignment descriptions of positions together with post-2013 data to predict which descriptions correspond to positions that would have been evaluated for reading and/or math effectiveness had the evaluation system been in place.

[9]Vacancy postings are restricted to teaching positions requiring more than a half–time schedule.

Table 1: Descriptive Statistics for Teachers & Open Positions

Panel A: Teacher Demographics	All Teachers	Teachers in ITS
	Mean or Percentage (SD)	
% Male	25.2	22.8
% Black	8.5	10.0
% Hispanic	3.3	2.9
% White	79.2	81.0
% Asian	4.4	4.3
Years of experience (mean)	13.4 (9.5)	11.2 (8.5)
Years of education (mean)	5.0 (1.0)	5.0 (1.1)
Years in sample (mean)	4.0 (2.8)	4.4 (2.4)
Teacher count	823	421
Teacher–year count	3,268	1,861
Panel B: ITS – Teacher Applicants		Teachers in ITS
		Mean (SD)
Size of position choice menu (mean)		36.5 (15.6)
Applications submitted (mean)		6.7 (7.1)
Number of interviews (mean)		3.4 (3.1)
Number of times ranked (mean)		1.5 (1.6)
Number of offers (mean)		0.8 (1.1)
Panel C: ITS – Open Positions		Positions in ITS
		Mean (SD)
Number of potential applicants (mean)		213.7 (87.4)
Number of applicants (mean)		4.8 (3.9)
Number of interviews (mean)		2.5 (1.9)
Number of offers (mean)		0.6 (0.7)
Position count		972

Note: Panel A shows the percentage (%), mean, counts, and standard deviation (SD) of teacher characteristic for all the teachers in the sample in the column on the left and for all the teachers who ever applied to an open position in the Internal Transfer System (ITS) in the column on the right. For teachers who are ever found in the ITS, i.e., teacher applicants, we average their characteristics for every year they are in the sample between 2010 and 2019, including for those years they are not found in the ITS. Panel B shows the mean and SD of additional characteristics for teachers who applied to an open position in the sample, while Panel C shows the mean and SD of characteristics and the count of open positions in the ITS.

mandated standardized test scores in math and reading and their grade, school, demographic characteristics such as race, ethnicity, and gender, indicators of English language learner status, special education status, and eligibility for free or reduced-price lunch

(FRPL), which proxies for low household income. The sample includes 35,608 students in grades 4 through 8 in 54 schools between 2009 and 2019. Approximately a third of the students are Black, and a third are White. Hispanic students constitute approximately 20 percent of the sample. Last, for each school–year pair, we observe the characteristics of all teachers, and students at a school that potential hires may consider when making their decisions.[10]

2.3 Descriptive Evidence

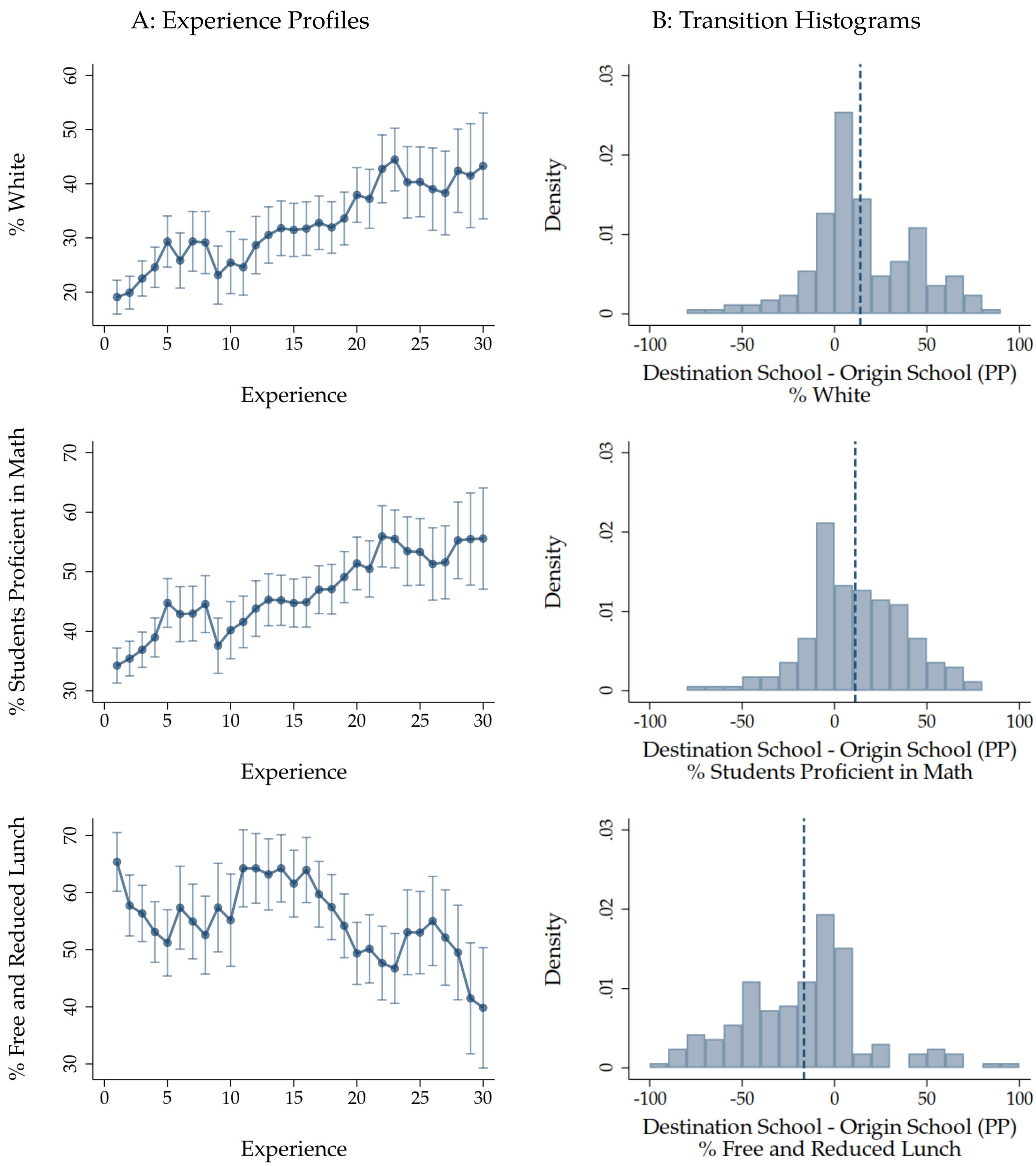
To motivate our model and counterfactual exercises, we begin by examining the observed assignment of teachers to schools and the transfer patterns of teachers over their careers. Teachers with more experience tend to serve in positions at schools with a higher share of White students, a greater share of students proficient in mathematics, and a lower share of low-income students (Figure 1: Panel A). Teachers' decisions during the transfer process partially account for the patterns observed in this figure, although other forces—such as initial assignments and differential retirement or exit rates—may also contribute.

Focusing on teachers who transferred schools between 2010 and 2019, we observe that teachers tend to move to schools with a higher share of White students, a higher share of students proficient in mathematics, and a lower share of low-income students than those of their prior school (Figure 1: Panel B), although there is considerable heterogeneity across these measures. This finding is consistent with existing literature indicating teachers of more disadvantaged students are more likely to transfer to other schools (Boyd et al., 2005; Scafidi et al., 2007; Goldhaber et al., 2011, 2019; Isenberg et al., 2022).

The observed teacher-student assignment in Figure 1 arises from teachers' decisions in the transfer process, which determine where they apply, accept offers, and ultimately work. These choices may be correlated with teachers' potential impact on students, though they need not place teachers in the classrooms where they are most effective. Accounting for such correlation is important both for assessing the potential gains from alternative assignments and for identifying teachers' effectiveness. On the one hand, the scope for improvement through alternative assignment depends on how much the

[10]See Appendix Table C.1 for additional school-level descriptive statistics.

Figure 1: Teacher Experience and Transition Profiles



Note: Panel A shows the mean and 95% confidence intervals of a school characteristic at the school at which teachers are assigned with each year of experience. Panel B restricts the sample to teachers who changed schools between 2010 and 2019, and it plots the histogram of the difference in school characteristics between the destination and origin schools. The vertical line is the mean difference. "PP" stands for percentage points.

observed matches depart from ideal placements. On the other hand, if teacher placements are correlated —even imperfectly— with their potential impact on students, separating teachers' general effectiveness from classroom–specific effectiveness becomes empirically challenging. To address this, we develop a model of teachers' transfer decisions and students' potential outcomes in Section 3 that allows for correlation between the two, capturing the possibility of selection into jobs. The model quantifies the extent to which teacher mobility patterns are related to student achievement gains.[11]

3. Model

We jointly model students' outcomes and teachers' labor market decisions to incorporate the potential for selection of teachers into schools.

3.1 Model of Student Outcomes

Equation 1 describes the learning outcome of a student-year pair k in subject m when assigned to teacher i at school s:

$$Y_{kism} = X'_{kis}\alpha + \theta_{im} + N'_{ik}\gamma + M'_k\beta_i + \eta^y_{is} + \varepsilon_{kism}, \tag{1}$$

where X_{kis} is a matrix with student, teacher, and classroom characteristics.[12] Characteristics include a student's past test scores and demographics, teacher demographics, education, and experience, class size, classroom averages of student demographics and average past test scores, and school by year fixed effects.[13] The term θ_{im}, which we refer to as *teacher general effectiveness*, captures teacher i's average impact on student outcomes in subject m, where $m \in \{\text{math}, \text{reading}\}$.

Match effects between students and teachers are represented by $N'_{ik}\gamma + M'_k\beta_i + \eta^y_{is}$, where N_{ik} is a matrix of interactions of teacher and student characteristics and γ is a vector of

[11]Beyond the role of helping us capture selection, modeling teacher decisions is essential in the counterfactuals as we'll consider the set of teacher acceptable assignments. A social planner replaces principals, so modeling them is not necessary.

[12]For our estimation, a "classroom" is a grade–school–year. Our data measure teacher-student matches at the grade–school–year level. We discuss this feature of our data in Section 4.1.

[13]Student demographics include their status as White non-Hispanic versus BIPOC, gender, free or reduced price lunch status, and English language learner and special education designations. Teacher demographics include their White non-Hispanic status and their gender.

slope coefficients on these interactions and is assumed common to all teachers. Interactions include same-race and same-gender terms, and the product of teacher experience and variables indicating whether a student is Black or Hispanic, whether a student is proficient (was proficient on the prior year's test), and the class size.[14] We include same-race and same-gender interactions to allow the possibility that these are relevant contributors to match effects (Dee, 2004, 2005). Interactions with teacher experience are motivated by evidence presented in Figure 1 that shows systematic sorting by experience in our data: allowing for match effects on teacher experience will allow us to capture selection on this margin, if it exists. For added flexibility, the term $M_k' \beta_i$ captures match effects of teacher i along student characteristics in M_k, which includes indicators for whether a student is Black or Hispanic and whether a student is proficient. The teacher-specific slope parameter β_i flexibly captures the effectiveness of teacher i with the relevant student type relative to the average teacher in the sample. These match effects are beyond those explained by a teacher's experience, race, or gender. Both $N_{ik}' \gamma$ and $M_k' \beta_i$ capture match effects that map into student observable characteristics; we refer to the sum of both as the *observable match effects*. Finally, we refer to η_{is}^y as the *unobservable match effect* of teacher i with students at school s. η_{is}^y captures the residual effectiveness of teacher i with all students at school s that is not rooted in observed student characteristics. This match effect reflects that some teachers may thrive with students of a specific unobservable type that are more prominent in school s, or in schools with certain characteristics. This may include, for example, the work climate or leadership style of the principal.

We make three assumptions about teacher productivity. First, we assume there are no complementarities in productivity across teacher peers. This means, a teacher's effectiveness at school s does not depend on the identity or the characteristics of other teachers at s. There is some evidence against this assumption (Jackson and Bruegmann, 2009), but it is standard in the literature and allows us to consider the alternative assignment of each teacher in isolation, without having to consider all potential teacher peer groups, which may lead to computational intractability when considering the alternative assignments of

[14]The class size is not a student characteristic, rather a classroom characteristic, but we leave it in N_{ik} for convenience. Classroom size the average for the school–grade–year, the number of such students divided by the number of full time equivalent (FTE) teachers assigned there.

teachers. Second, we assume teacher–school match effects vary in time only via changes in teacher experience and changes in student body composition. This rules out the possibility that, as teachers spend more time at a school, they gain school–specific knowledge that makes them more productive at that school. Third, we assume match effects are constant across subjects —math and reading— for each teacher, which implies our estimated match effects capture excess productivity that is present in student performance across both subjects. For teachers who teach both math and reading, almost 70% of our sample, our estimates of match effects are conservative as we require these to be consistent across both students' reading and math scores.

3.2 Model of Teachers' Decisions

We model teachers' decisions in the transfer process using a random utility framework. Equation 2 describes teacher i's utility from being assigned to position p, where each position corresponds to a specific school s and year t.[15]

$$U_{ip} = \overline{W}'_{st}\rho + \overline{N}'_{ist}\varphi + (\overline{M}_{st} \odot \beta_i)'\kappa + \lambda D_{ist} + \pi I_{ist} + \delta_s + \eta^u_{is} + \epsilon_{ip}, \tag{2}$$

The matrix $\overline{W}_{st}$ includes school-year averages of a subset of classroom demographics in X'_{kis}[16] and baseline test scores. The elements of ρ capture teacher preferences for these student characteristics. The vector $\overline{N}'_{ist}$ contains school–year averages of interactions in N_{ik}, which include the share of same–race and same–gender students for teacher i, interactions between i's experience and the share of Black and Hispanic students, the share of proficient students, and average class size in school s in year t. The parameters in φ capture heterogeneity in teacher preferences along these dimensions.

The term $(\overline{M}_{st} \odot \beta_i)'\kappa$, where $\odot$ denotes element–wise multiplication, captures interactions between school–level student composition and teacher–specific effectiveness.

[15]For purposes of estimation, a position is defined either as a job posting listed in the ITS or as a teacher's current assignment, which we treat as the outside option. Each job posting in our data is associated with a school, a year, and a grade, and includes a specific instructional assignment (e.g., "4th-grade math instruction").

[16]These demographics include student race or ethnicity, gender, free or reduced price lunch status, and English language learner and special education designations; we omit averages of teacher characteristics at the school–year level.

Here, $\overline{M}_{st}$ includes the average classroom shares of Black or Hispanic students and proficient students in each school-year, β_i measures teacher i's relative effectiveness with each group, and the parameters in κ govern how teachers' preferences for teaching these student groups vary with their match effects.

D_{ist} is the driving time between teacher i's home and a school s in year t. This driving time serves as a supply shifter and is excluded from the model of student outcomes. In Section 4, we discuss the assumptions and identification arguments behind this shifter in detail. η^u_{is} represents an idiosyncratic taste shock of teacher i for school s, and ϵ_{ip} is an idiosyncratic taste shock of teacher i for position p. The model also includes school fixed effects, in δ_s, capturing each school's unobserved attractiveness that is common across teachers over time.

Because teachers face the decision of whether to apply for a transfer within the school district each year, we include an inertia term in teacher utility that captures the cost to teachers of changing jobs. This includes the value teachers place on not having to update their CV, change their routines, form new networks and friendships, and adapt or generate new teaching material. We assume the value of inertia is the same for all teachers. The inertia cost is given by the parameter π. I_{ist} is an indicator variable that turns on for a teacher's current position.

Since all the parameters in Equation 2, except for ϵ_{ip}, are school specific, any within–school position–level disagreements will be captured by variation in the error term, ϵ_{ip}. These situations may capture instances in which, for example, a teacher applies to only one of two open positions at a school, indicating she finds one position more appealing than the other within the same school.

3.3 Parametric Assumptions and the Hierarchical Structure of the Model

We do not impose restrictions on the correlation structure of the unobservables across the two models, and we further assume $\theta_{im} \sim \mathcal{N}(0, \sigma^2_{\theta m})$, $\beta^1_i \sim \mathcal{N}(0, \sigma^2_{\beta 1})$, $\beta^2_i \sim \mathcal{N}(0, \sigma^2_{\beta 2})$, $\eta_{is} = (\eta^y_{is}, \eta^u_{is}) \sim \mathcal{N}(0, \Sigma_\eta)$. In addition, $\varepsilon_{kis} \sim \mathcal{N}(0, \sigma^2_{\varepsilon y})$, $\epsilon_{ip} \overset{iid}{\sim} \mathcal{N}(0, 1)$. The parameters to be estimated are the coefficients $\alpha, \gamma, \rho, \varphi, \kappa, \lambda, \pi$, and the variance–covariance matrices

$\sigma^2_{\theta m}$, and Σ_η, and the variance $\sigma^2_{\varepsilon y}$.

The hierarchical structure of the data—multiple students nested within each teacher—naturally induces shrinkage in our estimates of teacher value added. In our model, teacher effects θ_{im} are drawn from a Normal prior centered at zero, so the posterior estimates are pulled toward this prior mean. Because the idiosyncratic error variance $\sigma^2_{\varepsilon y}$ is assumed to be homoskedastic, its estimated value affects the overall degree of shrinkage: a larger $\hat{\sigma}^2_{\varepsilon y}$ implies stronger shrinkage toward zero. Differences in shrinkage across teachers arise through differences in the number of students they teach. Teachers with fewer student observations receive more shrinkage, while those with larger samples are shrunk less.[17] Similarly, estimates of teacher unobservable match effects (η^y_{is}) and individual–level match effects (β_i), are shrunk toward their prior mean of zero.

3.4 Sources of Selection

The parameters γ, φ, and κ capture teacher selection on match effects along student observable types. For example, the model of outcomes allows for a differential impact of teachers on students depending on whether both the teacher and student share the same race. This is captured in the parameter γ. At the same time, the model of teacher preferences allows teachers to prefer to teach at schools with larger shares of students who share a teacher's race; which would be captured by the parameter φ. Similarly, positive values for the terms in κ indicate positive selection over student characteristics in M_k.

On the unobserved side, we allow for correlations between the teacher–school match effects in student outcomes (η^y_{is}) and the idiosyncratic shocks in teacher decision models (η^u_{is}). A positive correlation between η^y_{is} and η^u_{is} would imply that teachers tend to value schools at which they have higher unobserved (at least for the econometrician) match

[17] The posterior mean of θ_{im} in our hierarchical model is equivalent to the empirical Bayes estimator commonly used in the teacher value-added literature (e.g., Kane and Staiger (2008); Chetty et al. (2014a)). Under the prior $\theta_i \sim N(0, \sigma^2_\theta)$ and homoskedastic idiosyncratic errors $\varepsilon_{kis} \sim N(0, \sigma^2_\varepsilon)$, the posterior mean takes the standard shrinkage form $E[\theta_i \mid \text{data}] = \lambda_i \, \hat{\theta}^{\text{raw}}_i \quad \text{where} \quad \lambda_i = \frac{\sigma^2_\theta}{\sigma^2_\theta + \sigma^2_\varepsilon / n_i}$. Here $\hat{\theta}^{\text{raw}}_i$ denotes the raw (within-teacher) average residual, and n_i is the number of student observations for teacher i. The shrinkage weight λ_i shows that posterior estimates are pulled toward the prior mean (zero) more strongly when sampling noise is large relative to true between-teacher variance, reproducing the empirical Bayes structure highlighted in the prior literature.

effects.[18]

4. Identification and Estimation

4.1 Mapping the Model to the Data

We estimate the model parameters using the decisions of teachers in the ITS and the test scores of students under the observed teacher–classroom matches. At the end of each academic year t, every teacher who had an assigned teaching position within the school district can decide to apply for a transfer. We observe the licenses of every teacher each year and the correct licenses required for every open position in the ITS. With this information, we create a choice menu of positions for each teacher. Because we observe 10 years of job posting data and two rounds of hiring each year, we build at most 20 menus per teacher and can observe the decisions in each case. On average, teachers' choice menus have 37 schools, and conditional on their applying to any open position at all, teachers apply to an average of 7 positions in each round (Table 1: Panel B).

Each year a number of teachers who had an assigned teaching position lose it. This can happen if enrollment at a school falls below a certain level, requiring a reduction in budget and staffing. If this occurs, the principal chooses the teacher who loses her assignment, which is called being "excessed". In principle, seniority protects teachers from losing their assignment, and more junior teachers are more at risk of being excessed. A teacher who loses her school assignment and wants to find a new one within the district goes to the ITS to search for a new position for the next academic year. Our data identify which ITS applicants had lost their prior assignment and which did not. In our sample, approximately 40% of the teachers searching for a match in the ITS each year had lost their assignment. For these teachers, the value of remaining unassigned after the ITS process ends is the expected value of the next round (only in the case of round 1), or the match they expect to find in the scramble round, in which every unmatched teacher must match with a remaining position. In contrast, a teacher who did not lose her assignment has her current position as her fallback option. Consistent with this part of the ITS process, we

[18] We do not take a stance on what teachers know of their match effects. Teachers may be making labor market decisions knowingly of their match effects, or could only imperfectly know about them.

assume that, in each round, a teacher applies to every position on her choice menu that she prefers to her fallback option, net of inertia. The inertia cost π applies only to teachers who did not lose their position (over half of the teachers in the ITS).

At the application stage, we assume that teachers consider all the positions on their choice menu. This means teachers are aware of all these positions and can compare them. The teaching positions in our sample follow standard descriptions and are differentiated mainly by the school and position type. Both are easily observable characteristics that a teacher can understand before coming to the ITS in a given year. We assume there are no application costs for teachers due to the infrastructure of the platform that collects applications. Applicants upload a standardized CV and can easily apply for multiple positions by clicking to apply to each position of interest, with the platform automatically distributing their materials to the respective hiring teams.

In the data corresponding to the stage at which teachers receive offers, we observe a subset of teachers who receive multiple concurrent offers and choose one. We assume teachers choose their preferred offer. We use this choice as an additional source of variation to estimate teacher preferences.

We now turn to the mapping between the data and the student outcomes component of the model. We link student test scores to teachers using observed assignments at the school–grade–year level and use job descriptions to classify each assignment as math instruction, reading instruction, or both. Because this matching is coarser than classroom-level roster data, it may introduce measurement error in student–teacher links and, in turn, attenuate estimates of teacher effectiveness and match effects. To assess the importance of this concern, Section 5.3 presents robustness estimates based on a subsample of teachers for whom classroom–level matches are observed. Estimates from this subsample are quantitatively similar to those from the full sample, suggesting that any bias arising from coarse assignment matching is limited and does not materially affect our main conclusions.

4.2 Estimation Strategy and Identification

We estimate the joint distribution of the coefficients $\alpha, \gamma, \rho, \varphi, \kappa, \lambda, \pi$, and the variance–covariance matrix Σ_η, and the variances $\sigma^2_{\theta m}$ and $\sigma^2_{\varepsilon y}$ using a Gibbs sampler and assuming conjugate uninformative priors (Gelman et al., 2013). Using this method, we generate draws of the joint distribution of the parameters and latent variables in the model. We draw 110,000 iterations of the sampler, burn 50,000 initial iterations, and keep only 1 of every 10 draws to reduce autocorrelation in the chains. We inspect the chains for convergence before reporting. Details on the estimation strategy can be found in Appendix B.

To identify the model of student potential outcomes, we require variation that allows us to separate teacher general effectiveness from unobserved teacher–school match effects. The central identification challenge is that teachers may sort systematically into schools with which they have high or low unobserved match productivity. Our strategy addresses this concern by exploiting an excluded shifter for teacher decisions that is unrelated to students' potential outcomes. By introducing quasi-random variation into teachers' assignment decisions, the shifter identifies the dispersion of unobserved match effects, $\text{Var}(\eta^y_{is})$, using cross–teacher variation in school–specific effectiveness among teachers whose assignments are driven primarily by the shifter rather than by selection. Selection on unobservables is captured by the covariance $\text{Cov}(\eta^y_{is}, \eta^u_{is})$ and is identified by contrasting the student outcomes associated with teachers whose assignments are more strongly influenced by proximity (and hence closer to quasi-random) with those of teachers whose assignments are less constrained by distance and therefore more likely to reflect active selection. Under this framework, a finding that teachers living closer to a school exhibit systematically lower match effects than those living farther away is interpreted as evidence of positive selection into schools based on unobserved match quality.

Driving Time Shifter. We use the driving time in minutes from a teacher's residence to each school in her choice set as a shifter of teacher utility, D_{ist}.[19] Driving time is expected to be relevant because positions closer to a teacher's residence entail lower commuting costs and are therefore more attractive. Consistent with this intuition, we show below that

[19]The mean driving time between teacher residences and open positions in our sample is 19 minutes, well below the 2017 U.S. average commute of 27 minutes (2018 American Community Survey, 1-year estimates).

shorter driving times increase the probability that a teacher applies to a given position and, conditional on application, increase the probability that the teacher and school ultimately form a match. This pattern is consistent with survey-based evidence indicating that teachers place substantial weight on commuting distance (Engel et al., 2014).

For the shifter to be valid, we require that driving time affects student outcomes only through its effect on the *probability* of assignment of teachers to classrooms. This requirement has two components. First, conditional on the realized teacher–school assignment, driving time does not directly affect student potential outcomes (*exclusion*), ruling out channels such as commuting fatigue or time costs that would alter teaching effectiveness within a given match. Second, conditional on observed determinants of teachers' choices –their preferences and constraints–, variation in driving time generates quasi-random variation in assignment probabilities: among teachers who are observationally similar and face the same feasible set of positions, living closer to one school rather than another does not proxy for teacher general effectiveness or unobserved teacher–school match effects (*conditional independence*). This assumption would be violated if, even after conditioning on observed determinants of teachers' preferences and choice sets, driving time systematically proxies for unobserved teacher quality or unobserved teacher–school match productivity–for example, due to residential location choices being correlated with effectiveness.

We use data on teachers' application behavior and the resulting matches in the ITS to assess the relevance of our distance shifter. Considering all teachers who applied to an open position and all positions they were eligible to apply to, we find a negative and significant relationship between driving time and the probability of submitting an application (Table 2: Column 1). The coefficient of -0.015 implies that a one-standard deviation reduction in driving time—approximately 18 minutes—is associated with a 11% increase in the probability of applying to a given job, relative to the sample's average application probability (0.137).

Driving time also predicts the probability of matching with a job. Using the same set of feasible teacher–position pairs, we find that a one-standard-deviation reduction in driving time is associated with a 21% increase in the probability that a teacher and a school form a match in the ITS (Column 2), relative to the sample's average match probability (0.003).

Table 2: Relevance and Falsification Tests for the Driving Time Shifter

	Relevance of Driving Time Shifter		Validity of Driving Time Shifter	
	Prob. of Application	Prob. of Matching	Prob. of Interview | App	Rank | Interview
	(1)	(2)	(3)	(4)
Drive time	-0.0151	-0.0007	-0.0100	0.0310
	(0.0018)	(0.0003)	(0.0070)	(0.0280)
Constant	0.1366	0.0033	0.4800	1.9900
	(0.0019)	(0.0003)	(0.0070)	(0.0310)
No. of observations	33,805	33,805	4,633	1,075

Note: This table shows regression results of the standardized (mean = 0 & standard deviation (SD) = 1) drive time (in minutes) between an open position and a teacher's home. Column 1 indicates the probability (Prob.) of application, while column 2 indicates the probability of a match between a teacher and the open position. The data is restricted to teachers who applied to at least one open job and to all the open positions each teacher could have applied for. Columns 3 and 4 show regression results of the drive time between an open position and a teacher, where column 3 is the probability of interview for candidates who applied to an open position, while column 4 is the ranking given to an interviewed candidate, conditional on them being interviewed.

While the exclusion and conditional independence requirements cannot be tested directly, we provide two falsification exercises, based on schools' interview and ranking decisions in the ITS, to assess the plausibility of these assumptions. If driving time were systematically correlated with match effects or general effectiveness, and if principals can partially anticipate or learn about such attributes during the hiring process, then their decisions of whom to interview and how to rank interviewed candidates would be expected to reflect that correlation. While principals may not perfectly observe teacher–school match quality, interview and ranking decisions represent the stage of the hiring process at which such information is most likely to be revealed, making these outcomes a natural setting to assess whether driving time proxies for productivity–related attributes. We find little evidence consistent with this concern. Among applicants for a given job, principals' interview decisions are not systematically related to teachers' driving times (Table 2: Column 3). Moreover, among interviewed candidates—where principals' information about match quality should be strongest—driving time does not predict the ranking principals assign in the ITS (Column 4). Together, these patterns provide indirect evidence that driving time operates primarily through teachers' job preferences rather than through dimensions of productivity observed by schools.

A potential violation of exogeneity would arise if teachers are systematically more effective at teaching students with whom they share a cultural background and are more likely to reside in neighborhoods with these families in a way not fully accounted for by race interactions. To address this concern, we incorporate a school attendance boundary dummy into our outcomes model with an indicator that identifies whether each school serves the teacher's neighborhood of residence.[20] Since the schools attended by the children of teachers are likely to serve students who share unobservable cultural traits with those teachers, the coefficient on this indicator variable will reflect the impact of cultural familiarity on student test scores. We find this additional control to be uncorrelated with student outcomes, suggesting little role of cultural proximity —as measured by attendance boundaries— on outcomes. We keep the control in our main specification to be conservative.

The use of an excluded shifter builds on a large literature on selection models that jointly model outcomes and choices to estimate treatment effects corrected for selection bias (Geweke et al., 2003; Heckman and Navarro, 2007; Lewbel, 2007; Hull, 2020). More closely, our strategy relates to the work by Walters (2018), Agarwal et al. (2025), and Kapor et al. (2024) who use excluded shifters to evaluate outcomes from assignment mechanisms.

5. Results

5.1 Parameter Estimates

In this section, we present the mean and 95% credible intervals of the posterior empirical distribution[21] of the parameters in the models of student outcomes and teacher preferences as shown in Tables 3 and 4.[22]

[20]This indicator is included in the matrix X_{kis}.

[21]The 95% credible interval is the interval between the 2.5th and 97.5th percentiles of the posterior empirical distribution.

[22]Our estimation derives the joint empirical distribution of the model parameters. We provide the mean and credible intervals of the marginal distribution for each parameter. In contrast to the frequentist approach,

Table 3: Estimated Parameters of the Student Outcomes Model

	Estimate	*95% credible interval*
Panel A: Teacher General Effectiveness		
Std. Dev. of Teacher General Effectiveness in math (θ_i^m)	0.101	[0.093,0.11]
Std. Dev. of Teacher General Effectiveness in reading (θ_i^r)	0.044	[0.034,0.054]
Panel B: Observable Match Effects		
Std. Dev. of effectiveness with Black or Hispanic students (β_i^1)	0.024	[0.02,0.029]
Std. Dev. of effectiveness with proficient students (β_i^2)	0.110	[0.101,0.12]
Same race	0.004	[-0.003,0.012]
Same gender	0.007	[0.004,0.01]
Teacher experience * Student Black or Hispanic	0.003	[-0.001,0.008]
Teacher experience * Student proficiency	-0.01	[-0.018,-0.002]
Teacher experience * Class size	-0.003	[-0.006,0.001]
Panel C: Unobservable Match Effects		
Std. Dev. of teacher–school Unobservable Match Effectiveness (η_{is}^y)	0.060	[0.054,0.067]
Panel D: Teacher Characteristics		
Male	-0.016	[-0.031,-0.001]
Education	-0.003	[-0.009,0.002]
Race - Black	0.012	[-0.009,0.033]
Race - Hispanic	0.017	[-0.013,0.044]
Race - other	-0.002	[-0.023,0.018]
Experience 2 to 3	0.003	[-0.009,0.014]
Experience 4 to 6	0.012	[-0.001,0.025]
Experience 7+	0.015	[0,0.03]
Catchment school dummy	-0.003	[-0.025,0.018]
Panel E: Student Characteristics		
Previous score	0.805	[0.801,0.808]
Previous score sq	0.005	[0.004,0.006]
Previous score cube	-0.004	[-0.004,-0.003]
Race - Black	-0.086	[-0.094,-0.078]
Race - Hispanic	-0.046	[-0.055,-0.038]
Race - other	-0.037	[-0.046,-0.028]
Male	-0.011	[-0.015,-0.008]
Low income	-0.069	[-0.073,-0.064]
English language learner	-0.012	[-0.017,-0.007]
Special education	-0.13	[-0.135,-0.125]
Panel F: Classroom Characteristics		
Share low income	-0.077	[-0.148,-0.006]
Share English language learners	-0.268	[-0.332,-0.207]
Share special education	-0.164	[-0.226,-0.102]

Continued on next page

which emphasizes statistical significance, presenting the mean and credible intervals allows us to visualize the complete distribution of the posterior.

Table 3: Estimated Parameters of the Student Outcomes Model *(continued)*

	Estimate	95% *credible interval*
Share Black	0.013	[-0.061,0.092]
Share Hispanic	0.141	[0.05,0.233]
Share other	0.145	[0.054,0.239]
Average previous score	-0.205	[-0.228,-0.182]
Average previous score sq	0.017	[-0.004,0.038]
Average class size	-0.004	[-0.009,0.001]

Note: The table shows the mean or standard deviation (Std. Dev.) and the 95% credible intervals of the estimated chains of parameters of the student outcomes model (Equation 1). Panel A presents estimates of the standard deviation of teacher general effectiveness for student math and reading achievement. Panel B presents estimates of parameters governing teacher match effects rooted in student observables, sometimes in interaction with teacher observables. Panel C reports the standard deviation of teacher–school match effects on unobservables. Panel D reports on parameters estimates for teacher characteristics. For race/ethnicity indicators, the omitted category is non-Hispanic White. Panel E reports the individual student characteristics' parameter estimates. Panel F reports the classroom student population characteristics parameter estimates.

Student Outcomes. Table 3 presents estimates of the student outcomes model. We interpret a few of the parameters so readers can verify their understanding. We report SDs of θ_{im}, η^{y}_{is}, β^{1}_{i}, and β^{2}_{i}.

We estimate standard deviations of general teacher effectiveness of 0.10 in math and 0.04 in reading. In comparison, estimates of the dispersion of teacher value added from homogeneous-effects models typically range from 0.10 to 0.15 in math and from 0.05 to 0.15 in reading (Bacher-Hicks and Koedel, 2023). Our estimates lie near the lower end of these ranges, which is consistent with a framework that allows for teacher-student match effects: once heterogeneity in effectiveness across students is explicitly modeled, less of the remaining variation is attributed to general effectiveness alone.

Matching on observable characteristics matters along several dimensions. Teachers differ in their effectiveness depending on whether students are Black or Hispanic, as captured by the race- and ethnicity-specific match component β^{1}_{i}. The standard deviation of this aspect of observable match quality is about 0.02 student achievement standard deviations (Panel B). Teachers also differ substantially in their effectiveness with academically proficient students in the prior year, as captured by β^{2}_{i}. The standard deviation of this proficiency-based match component is much larger–approximately 0.11 student achievement standard deviations–implying considerably greater scope for gains from matching teachers to students based on prior achievement than on race or ethnicity.

The remaining parameters in Panel B capture additional dimensions of observable match quality, including interactions between student and teacher characteristics such as shared race or gender, as well as heterogeneity in effectiveness across student types that varies with teacher experience. Because these parameters capture different and potentially offsetting dimensions of matching, their aggregate importance cannot be inferred directly from Panel B and is instead summarized in Table 5. Finally, Panel C shows that the standard deviation of teacher–school unobservable match effectiveness η^{y}_{is} is approximately 0.06 student achievement standard deviations, indicating that unobserved match components are quantitatively important relative to several observable dimensions.

Teacher Utility. Table 4 reports estimates from the teacher utility model. The positive coefficient in the first row indicates that teachers whose match effects are stronger when teaching Black or Hispanic students also exhibit a stronger preference for assignments with such students. In contrast, the negative coefficient in the second row implies that teachers who are relatively more effective with academically proficient students are, on average, less attracted to teaching them. Teachers also display preferences for assignments with students of the same race.

Relative to less-experienced teachers, more-experienced teachers prefer schools with a smaller share of Black or Hispanic students and a larger share of academically proficient students (Panel A). Combined with the finding that more-experienced teachers are estimated to be more effective with non-proficient students (Table 3, Panel B), this pattern implies negative selection along the proficiency–experience margin. For teachers with no experience, the main-effect estimates in Panel D indicate a preference for schools with higher-achieving students and a smaller Hispanic student share, as well as indifference to class size. The interaction between experience and class size suggests that more-experienced teachers are more averse to larger classes, although the corresponding 95% credible interval includes both positive and negative values.

Table 4: Estimated Parameters of Teacher Utility

	Mean	*95% credible interval*
Panel A: Observable Match Effects		
$\beta_i^1 *$ Share Black or Hispanic	37.715	[30.208,46.347]
$\beta_i^2 *$ Share proficient	-10.894	[-12.575,-9.102]
Share same race	0.925	[0.735,1.118]
Share same gender	-0.74	[-1.771,0.261]
Teacher experience * Share Black or Hispanic	-0.311	[-0.439,-0.193]
Teacher experience * Share proficient	0.234	[0.097,0.375]
Teacher experience * Class size	-0.021	[-0.069,0.026]
Panel B: Unobservable Match Effects		
Std. Dev. of Teacher-School Unobservable Match Effects (η_{is}^u)	0.73	[0.69,0.771]
Panel C: Teacher Characteristics		
Catchment school dummy	0.235	[0.061,0.415]
Inertia	3.928	[3.842,4.016]
Panel D: School Characteristics - Students		
Share low income	0.052	[-0.078,0.179]
Share English language learners	0.11	[-0.577,0.764]
Share special education	0.786	[-0.264,1.768]
Share Black	0.000	[-0.903,0.901]
Share Hispanic	-1.483	[-2.488,-0.53]
Share other	2.853	[1.661,4.093]
Average scores	0.33	[0.03,0.633]
Average scores sq	-0.277	[-0.577,0.018]
Average class size	0.037	[-0.021,0.095]
Panel E: Shifter		
Drive minutes	-0.248	[-0.299,-0.201]

Note: The table shows the means or standard deviations (Std. Dev.) and the 95% credible intervals of the estimated chains of parameter estimates of the teacher utility model (Equation 2). Panels A presents parameters associated with observable individual student characteristics constituting observable match effects. Panel B presents parameter estimates associated teacher–school unobservable match effects. Panel C reports on parameter estimates associated with observed teacher–school characteristics and Panel D with schools' student population characteristics.

As expected, teachers dislike assignments that require longer commutes, as reflected by the negative coefficient on driving time, which serves as the teacher utility shifter. This preference for shorter commutes is conditional on whether the teacher resides within a school's catchment area. Teachers' idiosyncratic tastes for schools, captured by η_{is}^u, play a substantial role in shaping assignment preferences, indicating considerable heterogeneity in school valuations even after accounting for observable characteristics. Finally, the estimates point to nontrivial inertia costs, consistent with teachers facing adjustment costs that make year-to-year reassignment less attractive, even when transfers are feasible.

5.2 Impact of Teacher Effectiveness and Match Effects

To quantify the relative contributions of teacher general effectiveness (θ_i), observable match effects ($N'_{ik}\gamma + M'_k\beta_i$), and unobservable match effects (η^y_{is}) to student achievement—and to characterize how these components relate to teacher decisions—Table 5 reports two sets of estimates. Column 1 presents the average change in student test scores resulting from an independent one SD change in each component of teacher quality across all feasible teacher assignments. These estimates capture the impact of reassigning an average student to a teacher with higher effectiveness, either via higher general effectiveness or match effects, considering all feasible alternative assignments.[23] Column 2 presents the corresponding change in teacher utility.

Raising students' assigned teacher math general effectiveness by 1 SD would lead to an average improvement in math student test scores of 0.12 SD, and of 0.05 SD in reading test scores when raising a teacher's reading general effectiveness (Table 5: Column 1). On observable match effects, a one SD increase would improve the average student test score by 0.08 SD across both math and reading. This aggregates the effects of all variables in Panel B of Table 3, such as whether a student is matched with a teacher of the same (or different) gender and race/ethnicity, and of proficient students, BIPOC students, and students in classes of different sizes are matched with more (or less) experienced teachers. Regarding the unobservable teacher–school match, the impact of being assigned a teacher

[23]Estimates of the impact of a one-standard deviation increase in teacher general effectiveness in math and reading are computed as the ratio of the estimated standard deviation of general effectiveness in each subject to the standard deviation of student test scores, where the latter is evaluated over a random sample of feasible teacher-student assignments. Analogously, the impact of unobservable match effects is computed as the ratio of the estimated standard deviation of the unobservable match component to the standard deviation of student test scores, using the same random sample and pooling across math and reading. Finally, the impact of observable match effects is computed as the ratio of the standard deviation of student-level observable match effects–calculated over the same random sample of feasible assignments–to the standard deviation of student test scores in that sample. The resulting distribution of observable match effects reflects the joint distribution of relevant student and teacher characteristics in the feasible set of teachers for each student.

Table 5: Impact of Teacher Effectiveness and Match Effects

	Student Outcomes Δy_{kt} (1 SD $\uparrow$) [95% *confidence interval*]	*Teacher Utility* Δu_{ist} (1 SD $\uparrow$) [95% *confidence interval*]
One standard deviation increase		
Teacher math general effectiveness, θ_i^m	0.12	
	[0.11,0.13]	
Teacher reading general effectiveness, θ_i^r	0.05	
	[0.04,0.06]	
Observable match effects	0.08	-0.11
	[0.07,0.08]	[-0.17,-0.07]
Unobservable match effects, η_{is}^y	0.07	-0.03
	[0.06,0.08]	[-0.07,0.03]

Note: The table shows the results from simulated changes in student achievement outcomes and teacher utility with a 1 standard deviation (SD) increase in teacher general effectiveness (θ_i), teacher match effectiveness from observables, and teacher–school match effectiveness from unobservables (η_{is}^y). Bootstrapped 95% confidence intervals are shown in square brackets and are built using a random sample of realizations of the estimated parameters, drawn from their empirical joint distribution.

with 1 SD higher match effects would result in a gain of 0.07 SD in student outcomes across both math and reading. All these differ statistically from zero across simulations.

Moving to the effects of general effectiveness and match effects on teacher utility, teachers are averse to positions in which they would have higher match effectiveness (Column 2). This is especially true for the portion of match effectiveness captured by observable characteristics. The negative selection estimate of -0.11 masks significant preference heterogeneity. As discussed above regarding Panel A of Table 4, teachers who have higher match effectiveness in teaching Black or Hispanic students also have higher taste for teaching them. But, those with greater productivity match for teaching proficient students have stronger aversion to teaching them. Along unobservable dimensions, the estimated coefficient of -0.03 points to negative selection as well, though the estimate is imprecisely measured.

5.3 Robustness

In this section, we assess the robustness of our results and provide additional validation and context. We address concerns arising from the absence of classroom-level roster data by re-estimating the model on an alternative sample, corroborate our teacher effectiveness measures using district-provided metrics, and examine the external validity of our sample.

Re-estimating the model in single-teacher classrooms: The district maintains a roster-verification process in which teachers verify their own student rosters, and these verified rosters are used to compute the district's value-added measures. Our data, by contrast, identify teacher assignments at the school-grade-year (and subject) level. This coarser matching raises the concern that estimates of the variance of teacher general effectiveness or match effects may be attenuated, or otherwise biased, due to measurement error in student-teacher matches.

To address this concern, we re-estimate the model using the subsample of school-grade-year-subject cells that contain only a single teacher. In this subsample, pooling is eliminated by construction, and the teacher–student match is effectively observed, substantially reducing concerns about measurement error. A little over 32 percent of school-grade-year-subject cells in our sample meet this criterion.

Single-teacher cells are more common in smaller schools and are associated with slightly larger class sizes and lower shares of English language learners. Importantly, however, these classrooms have comparable baseline achievement levels, racial composition, and shares of low-income students relative to multi-teacher cells. Teachers in single-teacher classrooms are more likely to be White but have similar levels of experience. Mean differences across subsamples are generally small, and when statistically detectable, do not suggest economically meaningful divergence.

Table C.5 reports estimates for this subsample. Consistent with attenuation arising from coarser matching, the estimated standard deviation of teacher general effectiveness increases modestly–from 0.101 to 0.117 in math and from 0.044 to 0.065 in reading–while the standard deviation of teacher–school unobservable match effects is essentially unchanged (0.060 to 0.059). For observable match effects, the standard deviation of effectiveness with Black or Hispanic students increases from 0.024 to 0.040, and with proficient students from 0.110 to 0.127. Overall, these changes are modest, suggesting that any attenuation induced by school-grade-year matching is limited and does not materially affect our main conclusions.

Comparing our value-added to the district's roster-based value-added: We compare our estimates of teacher general effectiveness to the district's value-added measures, which are computed using detailed classroom roster data and a homogeneous value-added model. Appendix Table C.2 shows that our standardized math effectiveness measure is strongly positively correlated with the district's math value-added (0.499), and also our standardized reading effectiveness is positively correlated with the district's reading value-added (0.254).[24] These correlations provide reassurance regarding our modeling choices and estimation methodology.

Correlations with other district effectiveness measures: We also compare our estimates of teacher general effectiveness to two additional district evaluation measures that are not based on test-score growth: a student survey–based measure constructed from teachers' own students using roster-verified matches, and an observation-based score derived from repeated classroom evaluations conducted by certified peer raters using a standardized instructional rubric. Because these measures capture dimensions of teaching practice and classroom experience that need not map one-to-one into achievement growth, we interpret the resulting correlations as descriptive evidence that our estimates relate to independent indicators in sensible ways. Appendix Table C.2 shows positive correlations between our measure of math general effectiveness and both the survey-based evaluation (0.163) and the observation-based measure (0.156), with smaller corresponding associations for reading general effectiveness. These magnitudes are consistent with benchmarks from the MET project and related multiple-measures evidence, which typically finds that correlations between value-added and student surveys or observation-based measures are positive but modest, and larger in math than in reading (Kane and Staiger, 2012).

External Validity: The student and school characteristics of the district we study strongly resemble those of the population of U.S. urban schools. We use the Generalizer tool, specifically designed to quantify the degree of generalizability between a sample of studied K–12 schools and a target inference population of schools (Tipton and Miller, 2022). The

[24]District evaluation measures are available beginning in 2013. Accordingly, the correlations are estimated using observations from 2013 to 2019.

Generalizer tool uses propensity scores to measure the similarity between the sample and inference population, yielding a generalizability index value between 0 and 1. We compare the schools in our sample to 15,389 U.S. schools in the population inference sample in which 5th, 6th, 7th, or 8th grade is taught, that are in an urban locale, and that are not charter schools. Based on parameters such as school size, percentage of low-income students, percentage girls, percentage White, percentage Black, percentage Hispanic, percentage U.S. citizens, and median family income, our analysis implies a generalizability index value of 0.81, which the tool characterizes as a high level of generalizability.

6. Counterfactual Teacher Assignments

This section aims to quantify the improvements in student outcomes resulting from alternative teacher assignments across positions within the same school district.

To quantify the potential gains from reallocating teachers across positions, we focus on the set of grade 4-8 teaching slots observed during our 10-year study period for which math or reading test scores are available. This set includes positions that appeared in the ITS at any point, as well as positions that never entered the system because they were never vacant during the sample period.

In this section, we define a position as a grade-school-year teaching slot. In the data, these positions are indexed by the teacher who occupied them in the observed assignment. In the counterfactual assignment problem, however, different teachers can be reassigned to these same slots. Because positions are year-specific, if teacher i is observed teaching fourth-grade students in school s in year t, this constitutes one position; if the same teacher teaches the same grade in the same school in year $t+1$, this is treated as a distinct position. Positions may be associated with math scores, reading scores, or both. We denote the full set of positions by $\mathcal{P}$.

The pool of teachers eligible for assignment to positions in $\mathcal{P}$ in year t includes teachers occupying positions in $\mathcal{P}$ in year t, as well as teachers who applied to at least one open ITS position in year t but were not assigned to a position in $\mathcal{P}$ that year. We further restrict each teacher's feasible set of positions using licensure requirements, so that only teachers with the appropriate credentials (e.g., a math license) may be assigned to corresponding positions.

We simulate counterfactual assignments of teachers to alternative positions under two possible policy objectives. First, maximize average student test scores. Second, maximize the share of proficient students. Both objectives are widely used by school districts, parents, and policymakers to evaluate school district's performance. Our main results consider student performance in math. Results for maximizing reading are in Appendix Section D.[25] We study each policy objective under three counterfactual assignment scenarios: 1) *Unconstrained*, 2) *No quits*, and 3) *No quits + No layoffs*. These are explained in detail below.

First, we consider the counterfactual scenario in which teachers are assigned to positions to maximize each of the two policy objectives given the positions in $\mathcal{P}$ and the set of teachers described above without any additional constraints. We refer to this scenario as the *Unconstrained* counterfactual. However, some assignments generated under the unconstrained counterfactual may be unacceptable to a teacher without an adjustment in pay. Therefore, the second counterfactual scenario restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment based on our simulation of teacher utility using the teacher utility model's estimated parameters.[26] We refer to the second scenario as the *No quits* counterfactual because it guarantees that any teacher does not have lower utility in the alternative assignment than that under her observed assignment.[27] Under the first two counterfactual scenarios, the pool of teachers who are candidates for assignment is larger than the set of positions available because assignments can go to either: (a) the teachers observed to be assigned to a position in our sample, or (b) the teachers who unsuccessfully sought a new position in our sample through the ITS.[28]

[25]At the end of this section, we will provide a brief summary of the counterfactual results and compare them with the baseline results.

[26]Because the observed assignment is always feasible under this restriction, the set of solutions is not empty.

[27]Our constraint is likely stronger than a constraint that would guarantee no teacher quits her assignment, as we are not comparing candidate assignments to a teacher's reservation utility. This is a conservative constraint on what teachers find acceptable.

[28]Teachers not observed to be assigned to one of the positions in our sample may be assigned to other

The third counterfactual scenario further restricts the teacher candidates for alternative assignments to only those teachers observed assigned to a position in $\mathcal{P}$ in each year, such as group (a) as described in the previous paragraph. We refer to this third scenario as the *+ No layoffs* counterfactual. It has the same restriction as the second counterfactual, but further excludes the (b) group of teachers not observed in any assigned position in the data. In this counterfactual, the number of available teachers and available positions is the same. It does not allow a change in who is teaching. It simply allows reshuffling of the teachers observed teaching across positions in each year. The gains from this counterfactual are the maximum potential gains a school district can realize without changing pay or laying off teachers protected by the collective bargaining agreement.

We generate counterfactual assignments by solving a linear program that enforces one-to-one feasibility: each position must be filled by exactly one teacher, and each teacher can be assigned to at most one position. We construct the objective function by computing the average student test score (and share proficient) at every feasible teacher-position match, and selecting weights so that the linear program maximizes average student test scores rather than average classroom test scores. Solving an analogous program at the student-teacher level is computationally infeasible, as the number of potential student-teacher pairs is prohibitively large.[29]

We quantify the gains from alternative teacher assignments under the assumption that the joint distribution of model parameters is known, along with the values of θ_i and β_i for each teacher i. This implies knowledge of each teacher's general effectiveness and match effects associated with observable characteristics, while requiring only knowledge of the *distribution*-rather than the realization-of unobservable match effects. By contrast, implementing the fully optimal assignment would require knowledge of the realized unobservable match effects η^y_{is} for each teacher–school pair. We therefore interpret our counterfactual assignments as those chosen by a school district with perfect information

positions outside our sample within the school district. These are the positions either not associated with test scores or with less than 0.5 full-time equivalents required.

[29]We solve the following linear program: $\max_a \sum_{i,p} a_{ip} \cdot y_{ip}$ subject to $a_{ip}(1 - c_{ip}) = 0$, $\sum_i a_{ip} = 1$, and $\sum_p a_{ip} \leq 1$, where $p \in \mathcal{P}$ indexes positions and i indexes teachers. y_{ip} is a weighted average of the outcomes of students associated with p, when assigned to teacher i. $a_{ip} = 1$ if teacher i is assigned to position p, and $c_{ip} = 1$ if teacher i is feasible for position p.

about teachers' production technologies.

We do not incorporate teacher inertia costs into the counterfactual assignments, as these exercises compare alternative static assignments to the observed allocation rather than modeling transitions from prior positions.[30] It is also important to note that the counterfactual outcomes capture only contemporaneous gains. We do not model the effects of implementing an alternative assignment repeatedly over multiple years, which could generate compounded gains through dynamic feedback. Instead, for each year we consider a one-shot reassignment of teachers and evaluate the resulting one-year achievement gains. Although we use multiple years of data for estimation, the counterfactual results summarize a one-year gain, rather than gains accumulated over time.

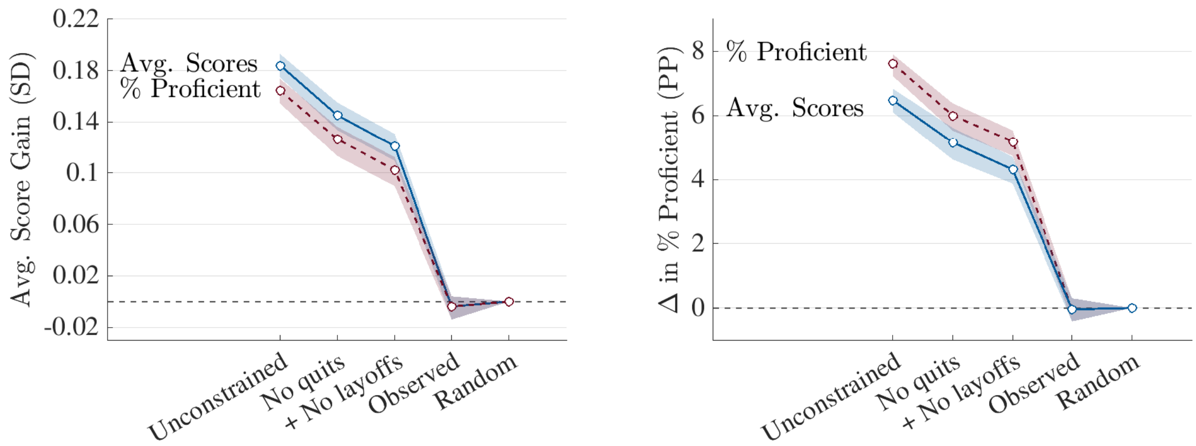
To implement these counterfactuals, for each scenario—including the observed and random assignments—we draw the model parameters 20 times from their estimated joint distribution. For each draw, we solve for the implied assignment and compute test score outcomes and the share of proficient students. We then report the mean and percentile distribution of gains across simulations.

6.1 Counterfactual Results

As seen in Figure 2a, under the first counterfactual scenario of *Unconstrained* assignment of teachers to positions, a policymaker seeking to maximize students' average math test score can push the average up by 19% of an SD relative to the test scores under the observed assignment (see Appendix Table C.3 for details). Under the second counterfactual scenario, where no retained teacher is harmed (*No quits*), average scores would increase by 15% of an SD relative to those under the observed assignment. Compared with the unconstrained assignment, this assignment is feasible without changing compensation, as retained teachers' welfare is not allowed to fall. In the third counterfactual scenario (+ *No layoffs*), where we restrict the teacher pool to include only those observed assigned positions in $\mathcal{P}$ and keep the *No quits* constraint, we still find substantial gains in test scores: 13% of an SD relative to those under the observed assignment. Unlike in this third

[30] Accounting for inertia costs is nevertheless important in estimation, as failing to do so may bias estimates of teacher preferences over school characteristics.

Figure 2: Counterfactual Gains in Math Test Scores and Percentage Proficient by Objective



(a) Outcome: Average Math Test Score

(b) Outcome: Percent Proficient in Math

Note: Panel (a) shows the gains in the average (avg.) student test scores in standard deviation (SD) terms and 95% confidence intervals under three counterfactual scenarios that aim to maximize the policy objectives of either the average math test score (solid blue line) or percentage proficient in math (dashed red line) relative to the outcomes under a random assignment. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual constrains each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The *+ No layoffs* counterfactual further constrains the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample. Panel (b) shows the gains in percentage points (PP) in math proficiency levels and 95% confidence intervals under the same two objectives and three counterfactuals.

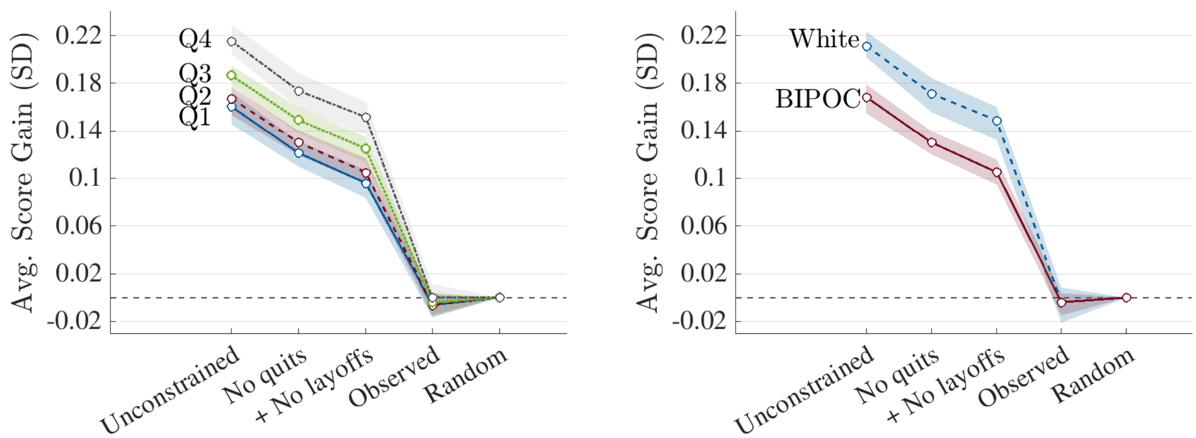
scenario, scenarios 1 and 2 can select observed unassigned teachers with higher general effectiveness to teach in the set of positions $\mathcal{P}$. Thus, these scenarios can and do result in higher gains in average student achievement.

When we consider a policymaker objective of maximizing the percentage of students proficient instead of average test scores, we find the percentage of proficient students would increase by 8 percentage points (PP) under an *Unconstrained* assignment relative to that under the observed assignment (Figure 2b).[31] Imposing the *No quits* constraint, a gain of 6 PP is possible. Further restricting the sample to only teachers observed assigned (*+ No layoffs*) reduces the gain to 5 PP relative to the share under the observed assignment. For completeness, the red line in Figure 2b shows the impact on percent proficient under each

[31] See Appendix Table C.4 for details.

counterfactual scenario when maximizing average achievement, rather than maximizing percent proficient. The red line in Figure 2a shows vice-versa.

Figure 3: Differential Gains in Average Math Test Scores by Achievement and Race Under the Policy Objective of Maximizing Average Math Achievement



(a) By Achievement Quartile

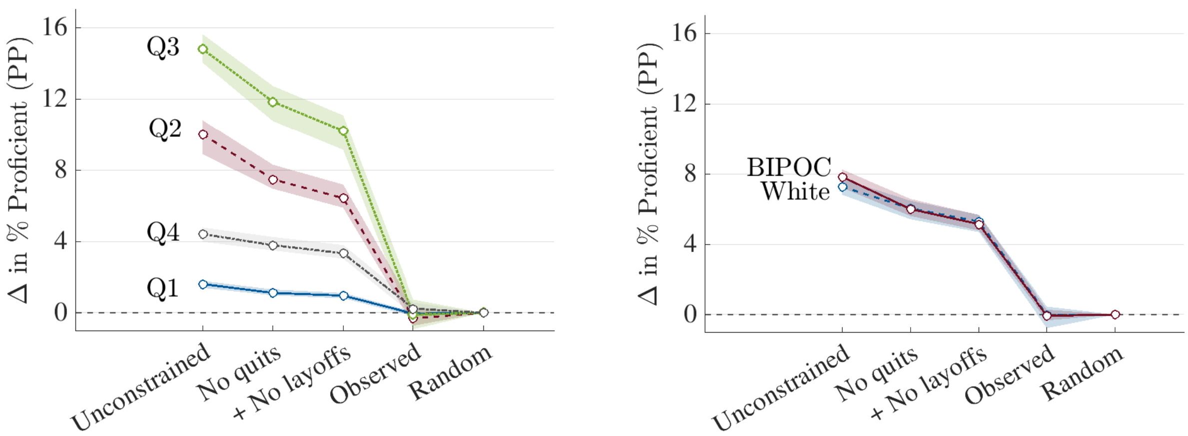
(b) By Race and Ethnicity

Note: Panels (a) and (b) show the average math test score gains in standard deviation (SD) terms and 95% credible intervals relative to those under the random assignment by baseline student achievement quartile (Q1-Q4) and student race and ethnicity, respectively, under the three counterfactual scenarios that maximize average math test scores. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The *+ No layoffs* counterfactual further restricts the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample.

Figure 3 shows that all three counterfactual assignments that aim to maximize the average student test scores disproportionately benefit higher-achieving (fourth quartile Q4 in Figure 3a) and White students (Figure 3b), even though student groups of all other types also experience average gains relative to their outcomes under the observed assignment of teachers to classrooms. All counterfactual assignments would widen the racial achievement gap and inequality in the overall achievement level in math in the school district, pointing to an equity–efficiency trade-off under all of these counterfactual assignments.

While the three counterfactual scenarios that maximize the average test scores benefit higher-achieving and White students more than other groups, the gains are less dispro-

Figure 4: Differential Gains in Math Percentage Proficient by Achievement and Race Under the Policy Objective of Maximizing Percentage with Proficient Achievement in Math



(a) By Achievement Quartile

(b) By Race and Ethnicity

Note: Panels (a) and (b) show the gains in percentage of proficient students in math in percentage points (PP) terms and 95% credible intervals relative to those under the observed and a random assignment by baseline student achievement quartile (Q1-Q4) and student race and ethnicity, respectively, under the three counterfactual scenarios that maximize the percentage of proficient students in math. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The + *No layoffs* counterfactual further restricts the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample.

portionate when the three scenarios maximize the percentage of proficient students in the district (Figure 4). These scenarios benefit students in the middle of the achievement distribution the most (quartiles 2 (Q2) and 3 (Q3)), while students in the two tails of the distribution experience lower gains. This is intuitive and consistent with findings by Neal and Schanzenbach (2010): classrooms with a larger share of students in the middle of the achievement distribution are close enough to the proficiency threshold that there are marginally more gains from an alternative assignment of teachers with higher general effectiveness toward them. In contrast, classrooms with many students in the top quartile are for the most part already proficient, and increase the percentage of proficient students in classrooms with many students in the bottom quartile would require alternative assignment that are too opportunity-costly. In addition, the gains for White and BIPOC students are similar across the three scenarios that maximize the percentage of proficient students, then this policy objective would not widen the racial student proficiency gap.

6.2 Counterfactual Decomposition and Discussion

To understand what drives the gains in each of the three scenarios, we decompose the gains from maximizing the average math score additively into the portions attributable to the observable and unobservable match effects and to teacher general effectiveness.[32]

Although match effects based on observables vary as much as unobservable match effects (as seen in Table 5), realizing those gains through assignment of teachers to classrooms is less feasible because each student's best teacher match is different and not every student can be assigned her best-matched teacher simultaneously without a change to the grouping of students across classrooms and schools.[33] In contrast, unobservable match effects are more easily realized as they are defined at the teacher–school level. Consequently, the decomposition in Table 6 shows that almost none of the potential gains in any of the scenarios is due to match effects on observables.

Better leveraging teacher general effectiveness contributes to potential achievement gains through two channels: keeping the least effective teachers out of classrooms (operates only in the *Unconstrained* and *No quits* scenarios) and matching the most effective teachers to larger classrooms. Across scenarios, these channels account for between a third and a quarter of the total potential gains. In contrast, unobservable match effects generate the bulk of the gains.

All these results focus on assignments to maximize average scores or percent proficient in math. Analogous exhibits in Appendix D show results for assignments that maximize reading. Results tend to be a bit smaller in the magnitude of potential gains, which is explained by the smaller relative magnitude of general effectiveness in reading compared to match effects in that case, but similar in shape. The main difference is that BIPOC students gain more than White students when maximizing reading proficiency, slightly

[32]Gain in share of students proficient can't be additively decomposed, so a decomposition excercise for this counterfactual is not presented.

[33]Similar results are found in Umosen (2024), who studies the role of classroom segregation in explaining gains from alternative teacher assignments.

Table 6: Decomposition of Gains when Maximizing Average Math Student Test Scores

	Decomposition of Gains Relative to the Observed Assignment (SD)			
	Total	*General*	*Match Effects*	
	Gain	*Effectiveness*	*Observable*	*Unobservable*
	(1)	(2)	(3)	(4)
Unconstrained	0.188	0.056	0.015	0.115
	[0.172,0.203]	[0.050,0.063]	[0.011,0.019]	[0.100,0.133]
No quits	0.149	0.048	0.010	0.088
	[0.136,0.160]	[0.042,0.055]	[0.007,0.014]	[0.077,0.099]
+ *No layoffs*	0.125	0.033	0.010	0.082
	[0.113,0.137]	[0.027,0.038]	[0.008,0.013]	[0.070,0.093]

Note: This table presents the decomposition of the gains in average math test scores in standard deviation (SD) terms under the three counterfactual scenarios relative to the outcomes under the observed assignment. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The + *No layoffs* counterfactual further restricts the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample. 95% confidence intervals in square brackets.

relaxing the equity-efficiency tradeoff in that case.

Before closing, consider that random assignment of teachers to students produces average student test scores and proficiency rates that are statistically indistinguishable from those under the observed assignment (Figure 2). At first glance, this may seem surprising, given that our model estimates negative Roy selection, which might suggest that the observed assignment would under-perform a random one. However, the result follows from the nature of the Roy selection. It operates primarily through observable match effects. Because classrooms are not perfectly stratified on observables, teachers cannot fully sort away from the classrooms where they would be most effective. As a consequence, the implied selection is negative but attenuated, yielding only small and statistically noisy differences between the observed and random assignments.

7. Conclusion

This paper studies the relative importance of teacher–student match effects and general teacher effectiveness in the production of student learning, and the implications of these components for teacher assignment policy. Using detailed administrative data from a large urban school district and a model that jointly captures teachers' transfer decisions and student outcomes, we exploit quasi-random variation in teacher assignments induced by differences in driving time between teachers and schools to isolate the role of match-specific productivity differences while accounting for non-random teacher sorting. This approach allows us to characterize the distribution of match effects, both observable and unobservable, and to quantify the learning gains attainable from alternative teacher assignments that respect teachers' preferences under existing compensation policies.

We find that match effects exhibit dispersion comparable in magnitude to general effectiveness, implying meaningful heterogeneity in teacher productivity across classrooms. Match effects associated with observable student characteristics and those arising from unobservable components have estimated impacts on student test scores of similar order, indicating that neither dimension dominates in explaining variation in student outcomes. In addition, we document systematic patterns of sorting: teachers tend to sort away from positions in which they are most productive along the proficiency margin, while sorting toward students with whom they are more effective along race and ethnicity. Taken together, these patterns imply overall negative Roy selection in observed teacher–position assignments.

We use the estimated model to evaluate counterfactual assignment policies. A fully informed planner who reallocates teachers to maximize average achievement, assuming no layoffs and subject to the constraint that no teacher assigned under the counterfactual is made worse off relative to her observed assignment, can raise average test scores by approximately 0.13 standard deviations. These gains are economically meaningful, corresponding to substantial increases in lifetime earnings, but they are unevenly distributed: higher-achieving and white students benefit more, leading to wider achievement gaps. Alternative objectives, such as maximizing the share of students who reach proficiency, yield slightly smaller average test score gains while mitigating these distributional consequences.

A decomposition of these gains shows that most of the improvement arises from better alignment along unobservable dimensions of fit rather than from matching on observable student characteristics. This pattern reflects the fact that classrooms typically enroll heterogeneous groups of students, which limits the extent to which observable match effects can be fully exploited through reassignment. By contrast, unobservable match components, which are captured at the teacher–school level in our model, offer a broader margin for improving teacher–student alignment under feasible assignment policies.

Taken together, these findings suggest that the potential gains from improved teacher assignment are comparable in magnitude to those estimated for aggressive personnel policies that focus on firing based on low effectiveness as measured by homogeneous value-added models. Prior work studying firing policies in settings that abstract from match effects finds that achieving gains of this size typically requires dismissing large shares of teachers, particularly among early-career cohorts (Staiger and Rockoff, 2010; Neal, 2011). For example, the policies analyzed in Staiger and Rockoff (2010) and Neal (2011) would remove roughly half of first-year teachers and smaller shares of more experienced teachers, yielding projected achievement gains on the order of one-tenth of a standard deviation. Accounting for the need to compensate teachers for increased firing risk further reduces the net benefits of such policies (Rothstein, 2015). By contrast, the gains we estimate arise from improved alignment between teachers and classrooms rather than from changes in the composition of the teaching workforce.

At the same time, the reassignment gains we estimate rely on strong informational assumptions about teacher productivity across contexts, particularly along unobservable dimensions of fit. They should therefore be interpreted not as a claim about the ease of implementing reassignment policies, but as a benchmark for the potential returns to institutions that can learn about and act on heterogeneity in teacher effectiveness across assignments. Because much of the scope for improvement operates along dimensions that are not directly observed, realizing these gains would require job structures and organizational practices that generate information about match quality. Exposure to a broader range of instructional environments–whether early in teachers' careers or later through roles that span schools–may allow districts to better infer where individual teachers are most productive, albeit at some up-front cost.

More broadly, this paper highlights that improving student achievement through teacher labor market policies need not rely solely on changes in compensation or dismissal rules. Instead, policies that improve how districts learn about and respond to heterogeneity in teacher effectiveness across classrooms represent a complementary margin for improving outcomes. By quantifying the potential gains from better assignment under realistic constraints, this paper reframes teacher reassignment not as an alternative to traditional personnel policies, but as an additional channel through which existing institutions may be leveraged to raise student achievement.

References

Abdulkadiroğlu, Atila, Parag Pathak, and Christopher Walters, "Who gets what may not matter: Understanding school match effects," Technical Report, Working Paper 2025.

Agarwal, Nikhil, Charles Hodgson, and Paulo Somaini, "Choices and outcomes in assignment mechanisms: The allocation of deceased donor kidneys," *Econometrica*, 2025.

Ahn, Tom, Esteban M. Aucejo, and Jonathan James, "The Importance of Student-Teacher Matching: A Multidimensional Value-Added Approach," *manuscript*, 2024.

Almagro, Milena and Aradhya Sood, "Sorting by Choice? Separating Preferences from Discrimination in Housing Markets," *manuscript*, 2025.

Angrist, Joshua, Guthrie Gray-Lobe, Clemence M Idoux, and Parag A Pathak, "Still worth the trip? school busing effects in boston and new york," Technical Report, National Bureau of Economic Research 2024.

Aucejo, Esteban, Patrick Coate, Jane Cooley Fruehwirth, Sean Kelly, and Zachary Mozenter, "Teacher effectiveness and classroom composition: Understanding match effects in the classroom," *The Economic Journal*, 2022.

Bacher-Hicks, Andrew and Cory Koedel, "Estimation and interpretation of teacher value added in research applications," *Handbook of the Economics of Education*, 2023, *6*, 93–134.

Bates, Michael, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin, "Teacher Labor Market Policy and the Theory of the Second Best*," *The Quarterly Journal of Economics*, 2024, p. qjae042.

Biasi, Barbara, Chao Fu, and John Stromme, "Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage," Technical Report, National Bureau of Economic Research 2022.

Bobba, Matteo, Tim Ederer, Gianmarco Leon-Ciliotta, Christopher Neilson, and Marco G Nieddu, “Teacher compensation and structural inequality: Evidence from centralized teacher school choice in Perú,” Technical Report, National Bureau of Economic Research 2024.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff, “Explaining the short careers of high-achieving teachers in schools with low-performing students,” *American economic review*, 2005, *95* (2), 166–171.

Campos, Christopher and Caitlin Kearns, “The Impact of Public School Choice: Evidence from Los AngelesâĂŹs Zones of Choice,” *The Quarterly Journal of Economics*, 2024, *139* (2), 1051–1093.

— , Jesse Bruhn, Eric Chyn, and Antonia Vazquez, “Who Chooses and Who Benefits? The Design of Public School Choice Systems,” Technical Report, National Bureau of Economic Research 2025.

Chetty, Raj, John N Friedman, and Jonah E Rockoff, “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates,” *American Economic Review*, 2014, *104* (9), 2593–2632.

— , — , and — , “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood,” *American economic review*, 2014, *104* (9), 2633–2679.

Combe, Julien, Olivier Tercieux, and Camille Terrier, “The design of teacher assignment: Theory and evidence,” *The Review of Economic Studies*, 2022, *89* (6), 3154–3222.

Condie, Scott, Lars Lefgren, and David Sims, “Teacher heterogeneity, value-added and education policy,” *Economics of Education Review*, 2014, *40*, 76–92.

Dee, Thomas S, “Teachers, race, and student achievement in a randomized experiment,” *Review of economics and statistics*, 2004, *86* (1), 195–210.

— , “A teacher like me: Does race, ethnicity, or gender matter?,” *American Economic Review*, 2005, *95* (2), 158–165.

Delgado, William, “Heterogeneous Teacher Effects, Comparative Advantage, and Match Quality,” Technical Report, Working Paper 2023.

Deming, David J, “Better schools, less crime?,” *The Quarterly Journal of Economics*, 2011, *126* (4), 2063–2115.

— , Justine S Hastings, Thomas J Kane, and Douglas O Staiger, “School choice, school quality, and postsecondary attainment,” *American Economic Review*, 2014, *104* (3), 991–1013.

Eastmond, Tanner S, Nathan J Mather, Michael David Ricks, and Julian Betts, “Welfare Added? Optimal Teacher Allocations with Value-Added Scores,” 2025.

Engel, Mimi, Brian A Jacob, and F Chris Curran, "New evidence on teacher labor supply," *American Educational Research Journal*, 2014, *51* (1), 36–72.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin, *Bayesian data analysis*, CRC press, 2013.

Geweke, John, Gautam Gowrisankaran, and Robert J Town, "Bayesian inference for hospital quality in a selection model," *Econometrica*, 2003, *71* (4), 1215–1238.

Goldhaber, Dan, Betheny Gross, and Daniel Player, "Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best?," *Journal of Policy Analysis and Management*, 2011, *30* (1), 57–87.

—, Vanessa Quince, and Roddy Theobald, "Teacher quality gaps in US public schools: Trends, sources, and implications," *Phi Delta Kappan*, 2019, *100* (8), 14–19.

Graham, Bryan S, Geert Ridder, Petra Thiemann, and Gema Zamarro, "Teacher-to-classroom assignment and student achievement," *Journal of Business & Economic Statistics*, 2023, *41* (4), 1328–1340.

Grissom, Jason A, Susanna Loeb, and Nathaniel A Nakashima, "Strategic involuntary teacher transfers and teacher performance: Examining equity and efficiency," *Journal of Policy Analysis and Management*, 2014, *33* (1), 112–140.

Hanushek, Eric A, "Education production functions," in "The economics of education," Elsevier, 2020, pp. 161–170.

— and Steven G Rivkin, "The distribution of teacher quality and implications for policy," *Annu. Rev. Econ.*, 2012, *4* (1), 131–157.

— et al., "Teacher deselection," *Creating a new teaching profession*, 2009, *168*, 172–173.

Heckman, James J and Salvador Navarro, "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 2007, *136* (2), 341–396.

Hull, Peter, "Estimating hospital quality with quasi-experimental data," *unpublished manuscript*, 2020.

Idoux, Clemence, *Integrating new york city schools: The role of admission criteria and family preferences*, MIT Department of Economics, 2022.

Irwin, Victoria, Ke Wang, Jiashan Jung, Elizabeth Kessler, Tolga Tezil, Safiyya Alhassani, Alexandra Filbey, Ryan Dilig, and Farrah Bullock Mann, "Report on the Condition of Education 2024," Technical Report NCES 2024-144, U.S. Department of Education, National Center for Education Statistics, Washington, DC 2024. Retrieved January 16, 2026.

Isenberg, Eric, Jeffrey Max, Philip Gleason, and Jonah Deutsch, "Do low-income students have equal access to effective teachers?," *Educational Evaluation and Policy Analysis*, 2022, *44* (2), 234–256.

Jackson, C Kirabo, "Match quality, worker productivity, and worker mobility: Direct evidence from teachers," *Review of Economics and Statistics*, 2013, *95* (4), 1096–1116.

—, "What do test scores miss? The importance of teacher effects on non-test score outcomes," *Journal of Political Economy*, 2018, *126* (5), 2072–2107.

— and Elias Bruegmann, "Teaching students and teaching each other: The importance of peer learning for teachers," *American Economic Journal: Applied Economics*, 2009, *1* (4), 85–108.

Kane, Thomas J and Douglas O Staiger, "Estimating teacher impacts on student achievement: An experimental evaluation," Technical Report, National Bureau of Economic Research 2008.

Kane, Thomas J. and Douglas O. Staiger, "Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains," January 2012. Measures of Effective Teaching (MET) Project.

Kapor, Adam, Mohit Karnani, and Christopher Neilson, "Aftermarket frictions and the cost of off-platform options in centralized assignment mechanisms," *Journal of Political Economy*, 2024, *132* (7), 2346–2395.

Laverde, Mariana, "Distance to schools and equal access in school choice systems," *Unpublished Manuscript*, 2024.

Lewbel, Arthur, "Endogenous selection or treatment model estimation," *Journal of Econometrics*, 2007, *141* (2), 777–806.

National Center on Teacher Quality, "Teacher Contract Database," https://www.nctq.org/contract-database/home 2022. Accessed: 2022-11-23.

Neal, Derek, "The design of performance pay in education," in "Handbook of the Economics of Education," Vol. 4, Elsevier, 2011, pp. 495–550.

— and Diane Whitmore Schanzenbach, "Left behind by design: Proficiency counts and test-based accountability," *The Review of Economics and Statistics*, 2010, *92* (2), 263–283.

Rothstein, Jesse, "Teacher quality policy when supply matters," *American Economic Review*, 2015, *105* (1), 100–130.

Roy, Andrew Donald, "Some thoughts on the distribution of earnings," *Oxford economic papers*, 1951, *3* (2), 135–146.

Scafidi, Benjamin, David L Sjoquist, and Todd R Stinebrickner, "Race, poverty, and teacher mobility," *Economics of education review*, 2007, *26* (2), 145–159.

Staiger, Douglas O and Jonah E Rockoff, "Searching for effective teachers with imperfect information," *Journal of Economic perspectives*, 2010, *24* (3), 97–118.

Tipton, E. and K. Miller, "Generalizer [Web-tool]," https://thegeneralizer.org 2022. Accessed: 2022-11-23.

Umosen, Ini, "Teacher Comparative Advantage, Achievement Gains, and Classroom Segregation," *manuscript*, 2024.

Walters, Christopher R, "The demand for effective charter schools," *Journal of Political Economy*, 2018, *126* (6), 2179–2223.

ONLINE APPENDIX

Match Effects and the Gains from Alternative Job Assignments: Evidence from a Teacher Labor Market

Mariana Laverde, Elton Mykerezi, Aaron Sojourner, & Aradhya Sood

A. Data Details

Teacher Data: The data on many teacher characteristics comes from publicly available personnel data from the state's Department of Education. The dataset is a teacher level panel containing each teacher-year's assignments, demographics, qualifications and credentials, and pay. Each assignment includes a code linking the teacher to specific grades at a specific school within the district and the role performed by the teacher (e.g., 5th grade general education teacher). A teacher assignment can be less than full time and each teacher can hold multiple assignments in a given year. We use this data to identify the teachers who are primarily responsible for students' math and reading instruction in a given year and tested grades (starting at grade 3). Teacher demographics include gender, race, ethnicity, and age. Teacher credentials include formal education and years of experience (which together largely determine their pay within a district) and state license information. The license data are used to determine which open positions each teacher would be eligible to apply for.

District's Internal Teacher Transfer System: Data on position postings and the teacher transfer process are proprietary administrative data provided by the district to us under the IRB protocol 1510S79046. The dataset consists of two files: an open positions file that lists each open position available to incumbent teachers (before they can be posted for external applicants) and a transaction file that documents each decision taken by a teacher candidate, a district, or a school official. The transaction file includes applications, invitations for interviews, post interview rankings, position offers, and offer acceptance/rejection decisions. The district also provided us a bridge file that links internal personnel data, including the teachers' home addresses, to the publicly available teacher data from the state's Department of Education.

Student Test Scores and Characteristics: Student test scores and characteristics are provided by the State's Department of Education as an anonymized panel at the individual student level that includes the entire history of their standardized test sores as well as demographic and socio-economic characteristics, including race, ethnicity, gender, a proxy for low family income (free- and reduced-price lunch or FRPL), English Language Learner (ELL), and special education status.

School Characteristics: The school's address is collected from publicly available data from the state's Department of Education. School×year average student and teacher characteristics are computed from their respective datasets. We compute shares of students in each ethnic and racial category, the share FRPL, ELL, and special education students, and the average class size.

Teacher Utility Shifter: The teacher utility shifter is the travel time between a teacher's home address and a school's address. To construct the shifter, we geocode teacher and school addresses and measure the driving times using the Google Maps API. We measure one way drive times on September 12, 2023 at 8 am.

School Attendance Zones Data: The school attendance zoning information comes from the School Attendance Boundary Survey conducted by the National Center for Education Statistics with assistance from the U.S. Census Bureau to collect school attendance boundaries for the 2015-2016 school year. We use the school-level attendance zone shapefiles to identify all the unique elementary, middle, and high school attendance boundaries in our school district. We then identify the assigned schools for all teachers whose home address is located within our school district by matching the geo-coordinates of the addresses to the school zones spatially.

B. Model and Estimation Details

The model consists of the following equations. The first one describes students outcomes Y_{kism}, and the second describes teacher utility U_{ip}:

$$Y_{kism} = X'_{kis}\alpha + \theta_{im} + N'_{ik}\gamma + M'_k\beta_i + \eta^y_{is} + \varepsilon_{kism} \tag{3}$$

$$U_{ip} = \overline{W}'_{st}\rho + \overline{N}'_{ist}\varphi + (\overline{M}'_{st}\beta_i)\kappa + \lambda D_{ist} + \pi I_{ist} + \delta_s + \eta^u_{is} + \epsilon_{ip} \tag{4}$$

We rewrite $(\eta^y_{is}, \eta^u_{is})$ as

$$\eta^y_{is} = f_{is,1}$$
$$\eta^u_{is} = \beta^u_1 f_{is,1} + f_{is,2}$$

where $f_{is,1} \sim N\left(0, \sigma_1^2\right)$, $f_{is,2} \sim N\left(0, \sigma_2^2\right)$.

Also, we assume, $\varepsilon_{kismt} \sim N(0, \sigma^2_{\varepsilon y})$, $\beta^1_i \sim N(0, \sigma^2_{\beta 1})$, $\beta^2_i \sim N(0, \sigma^2_{\beta 2})$, $\theta_{im} \sim N(0, \sigma^2_{\theta m})$.

The model parameters are:

$$\phi = \left(\alpha, \gamma, \rho, \varphi, \kappa, \lambda, \pi, (\delta_s)_s, \beta^u_1\right)$$
$$\sigma = \left(\sigma_1, \sigma_2, \sigma_{\theta m}, \sigma_{\varepsilon y}\sigma_{\beta 1}, \sigma_{\beta 2}\right)$$

Gibbs Sampler

Start with values of ϕ^0 and σ^0 from diffuse conjugate priors, and values f^0, β^0_i, u^0 for the latent variables. u^0, must be consistent with teacher application and offer acceptance decisions.

1. **Data Augmentation:** In this step, we update the values of the latent variable u^1 given the values of the parameters of the model, the rest of the realizations of the latent variables, and the application and acceptance decisions of teachers. We take the values of the remaining variables from the prior step.

2. **Update** $\alpha, \gamma, \rho, \varphi, \kappa, \lambda, \pi, (\delta_s)_s$

3. **Update** θ_{im} **and** θ_{ir}

4. **Update** β_i^1

5. **Update** β_i^2

6. **Update** $f_{is,1}$

7. **Update** $f_{is,2}$

7. **Update** $\sigma_1^2, \sigma_2^2, \sigma_{\varepsilon y}^2, \sigma_{\theta m}^2, \sigma_{\beta 1}^2, \sigma_{\beta 2}^2$: Note that $\sigma_{\theta m}^2$ represents two parameters, one for each subject. We restrict to θ's from each subject to estimate each.

C. Additional Results

Table C.1: Descriptive Statistics of Schools

	Mean	*Std. dev*	*p5*	*p95*
Share low-income	55.6	29.7	0.0	96.8
Share English language learner	23.8	19.3	1.7	61.2
Share Black	37.3	22.9	6.5	82.3
Share Hispanic	17.3	17.9	2.5	58.2
Share White	32.2	25.1	2.6	75.2
Share female	48.9	2.9	44.3	53.3
Share proficient - math	46.7	22.1	15.0	85.1
Share proficient - reading	46.9	22.8	13.6	84.5
Average teacher experience	13.7	4.0	8.0	21.0
Average teacher age	43.1	4.1	36.0	50.0
School-year observations	504	504	504	504

Note: The table shows the mean, standard deviation (Std. dev.), 5th percentile (*p5*), and the 95th percentile (*p95*) for various school level characteristics.

Table C.2: Relationships between Our Measure of Teacher Effectiveness and District Measures of Teacher Effectiveness

	St. Teacher Math Effectiveness (θ_i^m)			*St. Teacher Reading Effectiveness* (θ_i^r)		
Math VA	0.499					
	(0.043)					
Reading VA				0.254		
				(0.055)		
Survey-based Evaluation		0.163			-0.010	
		(0.054)			(0.056)	
Standards of Instruction			0.156			0.033
			(0.044)			(0.050)
Observations	563	574	615	545	558	595

Note: The table shows regressions where the dependent variable is the standardized estimated teacher effectiveness from our estimated model and the independent variables are four measures of teacher effectiveness built by the district. These include the district's own value-added (VA) measures constructed using classroom identifiers in math and reading, a measure based on student surveys, and a measure based on scores on an established rubric from four classroom observations by trained peer teachers each year. Standard errors are in parentheses.

Table C.3: Counterfactual Summary: Maximize Math Average Test Scores

	Average Test Score Gains Relative to the Observed Assignment (SD)		
	Unconstrained	No quits	+ No layoffs
All	0.188	0.149	0.125
	[0.172,0.203]	[0.136,0.16]	[0.113,0.137]
By Achievement			
First quartile	0.167	0.128	0.102
	[0.149,0.18]	[0.117,0.14]	[0.092,0.116]
Second quartile	0.173	0.136	0.111
	[0.156,0.187]	[0.124,0.148]	[0.1,0.125]
Third quartile	0.191	0.153	0.129
	[0.174,0.206]	[0.138,0.165]	[0.115,0.14]
Fourth quartile	0.216	0.174	0.152
	[0.196,0.234]	[0.157,0.185]	[0.136,0.165]
By Race			
BIPOC	0.172	0.134	0.109
	[0.155,0.185]	[0.123,0.147]	[0.098,0.123]
White	0.215	0.175	0.153
	[0.196,0.234]	[0.157,0.188]	[0.136,0.168]

Note: This table presents the average math test score gains relative to the observed assignment in standard deviation terms (SD) for all students as well as for students by baseline achievement (in quartiles) and race in a counterfactual where average student test scores are maximized. 95% confidence intervals in square brackets.

Table C.4: Counterfactual Summary: Maximize Math Percent Proficient

	Gains in Percent Proficient in Math Relative to the Observed Assignment (PP)		
	Unconstrained	No quits	+ No layoffs
All	0.077	0.06	0.052
	[0.071,0.082]	[0.056,0.066]	[0.048,0.058]
By Achievement			
First quartile	0.017	0.012	0.01
	[0.014,0.019]	[0.01,0.013]	[0.008,0.011]
Second quartile	0.103	0.078	0.067
	[0.09,0.112]	[0.071,0.087]	[0.06,0.076]
Third quartile	0.149	0.12	0.103
	[0.137,0.16]	[0.108,0.131]	[0.092,0.115]
Fourth quartile	0.042	0.035	0.031
	[0.039,0.044]	[0.033,0.038]	[0.028,0.033]
By Race			
BIPOC	0.079	0.061	0.052
	[0.072,0.085]	[0.057,0.068]	[0.048,0.059]
White	0.073	0.06	0.053
	[0.067,0.077]	[0.055,0.064]	[0.047,0.057]

Note: This table presents the gains in the percent of proficient students in math in percentage points (PP) terms relative to the observed assignment for all students as well as for students by baseline achievement and race, in a counterfactual where the percentage of proficient students is maximized. 95% confidence intervals in square brackets.

Table C.5: Estimates from the Student Outcomes Model - Single Teacher Classrooms

	Estimate	*95% credible interval*
Panel A: Teacher General Effectiveness		
Std. Dev. of Teacher General Effectiveness in math (θ_i^m)	0.1167	[0.093,0.141]
Std. Dev. of Teacher General Effectiveness in reading (θ_i^r)	0.0648	[0.04,0.092]
Panel B: Observable Match Effects		
Std. Dev. of effectiveness with Black and Hispanic students (β_i^1)	0.0401	[0.029,0.052]
Std. Dev. of effectiveness with proficient students (β_i^2)	0.1273	[0.109,0.148]
Same race	-0.021	[-0.043,0.001]
Same gender	0.02	[0.012,0.028]
Teacher experience*Student Black or Hispanic	0.006	[-0.007,0.018]
Teacher experience*Student proficiency	-0.033	[-0.053,-0.013]

Continued on next page

Table C.5: Estimated Parameters of the Student Outcomes Model *(continued)*

	Estimate	*95% credible interval*
Teacher experience* Class size	0.003	[-0.009,0.015]
Panel C: Unobservable Match Effects		
Std. Dev. of Teacher-School Unobservable Match Effectiveness (η^y_{is})	0.0589	[0.037,0.083]
Panel D: Teacher Characteristics		
Male	-0.021	[-0.057,0.016]
Education	-0.01	[-0.028,0.008]
Race - Black	0.019	[-0.044,0.081]
Race - Hispanic	0.02	[-0.119,0.157]
Race - Other	0.029	[-0.039,0.098]
Experience 2 to 3	0.009	[-0.028,0.047]
Experience 4 to 6	0.019	[-0.024,0.063]
Experience 7+	-0.002	[-0.048,0.045]
Catchment school dummy	-0.015	[-0.061,0.03]
Panel E: Student Characteristics		
Previous score	0.805	[0.797,0.814]
Previous score sq	-0.004	[-0.006,-0.002]
Previous score cube	-0.003	[-0.003,-0.003]
Race - Black	-0.111	[-0.134,-0.088]
Race - Hispanic	-0.066	[-0.092,-0.041]
Race - Other	-0.065	[-0.09,-0.04]
Male	-0.012	[-0.02,-0.004]
Low-income	-0.079	[-0.091,-0.068]
English language learner	-0.037	[-0.05,-0.024]
Special education	-0.102	[-0.115,-0.09]
Panel F: School Characteristics		
Share low income	0.027	[-0.214,0.264]
Share English language learners	0.17	[-0.069,0.409]
Share special education	-0.585	[-0.782,-0.392]
Share Black	-0.002	[-0.269,0.278]
Share Hispanic	0.026	[-0.325,0.372]
Share Other	-0.188	[-0.517,0.129]
Average previous score	-0.216	[-0.292,-0.137]
Average previous score sq	0.025	[-0.033,0.084]
Average class size	0.003	[-0.011,0.018]

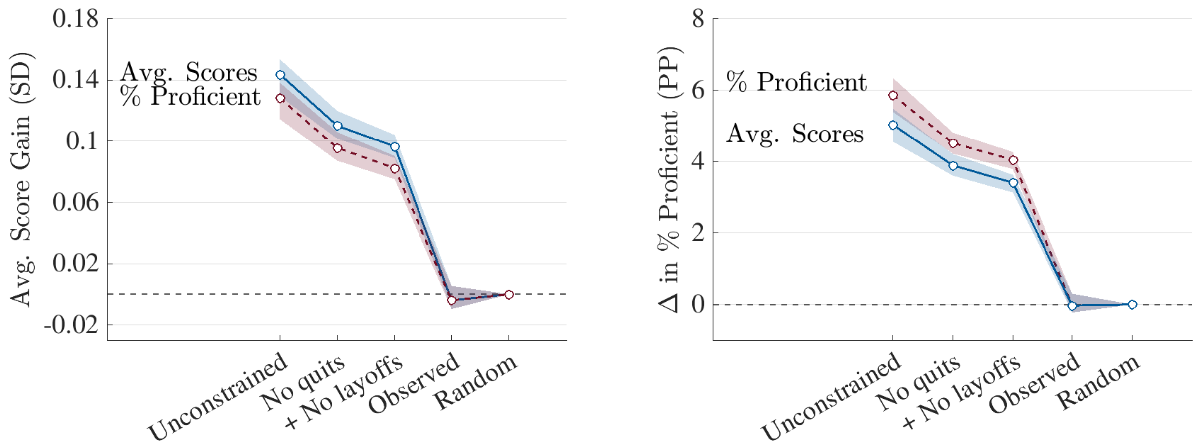
Note: The table shows the mean or standard deviation (Std. Dev.) and the 95% credible intervals of the estimated chains of parameters of the student outcomes model (Equation 1) when we restrict the sample to only single teacher classrooms. Please see the companion Table 3 for details on the estimated parameters.

D. Counterfactual Reading Results

Parallel to the counterfactual analysis of math achievement gains in the main text (Section 6), we examine counterfactual assignments chosen to maximize average reading achievement or the share of students proficient in reading. The overall patterns mirror those for math, though the magnitudes are somewhat smaller, consistent with the lower estimated variance of teacher general effectiveness in reading. Assignments that maximize average reading achievement raise scores by 0.148 SD relative to the observed assignment (Figure D.1a; Table D.1, top row). Imposing a constraint that no retained teacher is made worse off (*No quits*) reduces the gain to 0.114 SD, while additionally restricting assignments to teachers who are observed assigned to positions (+ *No layoffs*) still yields a gain of 0.101 SD. A decomposition of these gains highlights a dominant role for unobservable match effects (Table D.2), comparable to that found for math (Table 6). When the planner instead maximizes the share of students proficient in reading, the corresponding gains are 5.9, 4.5, and 4.1 percentage points across the three counterfactuals, respectively (Figure D.1b; Table D.3, top row).

A key difference relative to math counterfactual results is that the equity-efficiency tradeoff is considerably weaker for the reading results. Whereas maximizing math achievement disproportionately benefits White students, average gains from maximizing reading achievement are similar for White and BIPOC students (Figure D.2b). Moreover, when maximizing reading proficiency, estimated gains are larger for BIPOC students than for White students (Figure D.3b). One explanation is that BIPOC students are more heavily represented in the middle of the reading achievement distribution (quartiles Q2 and Q3) than in math, making proficiency-focused alternative assignments more equitable along the reading dimension.

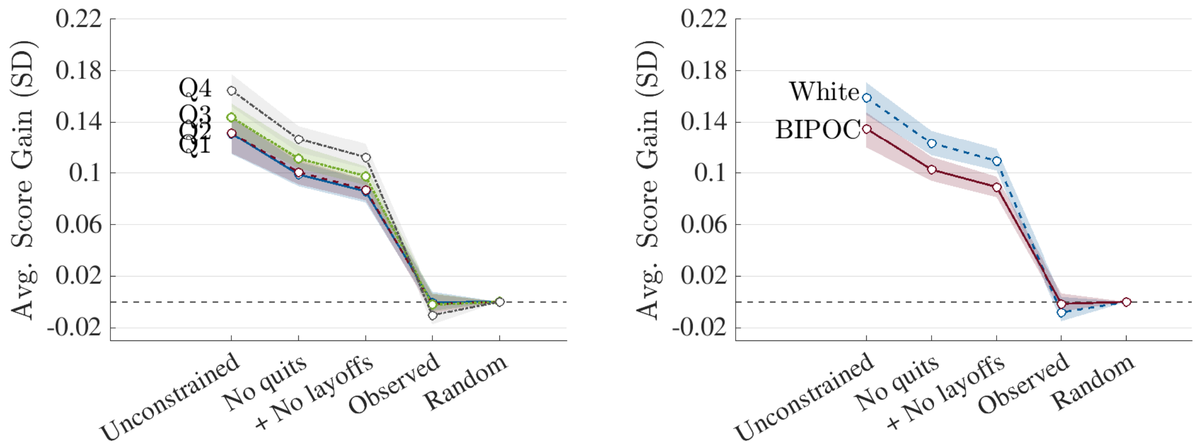
Figure D.1: Counterfactual Gains in Reading Test Scores and Percentage Proficient by Objective



(a) Objective: Average Reading Test Score (b) Objective: Percent Proficient in Reading

Note: Panel (a) shows the gains in the average (avg.) student test scores in standard deviation (SD) terms and 95% confidence intervals under three counterfactual scenarios that aim to maximize the policy objectives of either the average reading test scores (solid blue line) or percentage proficient in reading (dashed red line) relative to the outcomes under the observed assignment and a random assignment. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The + *No layoffs* counterfactual further restricts the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample. Panel (b) shows the gains in percentage points (PP) in math proficiency levels and 95% confidence intervals under the same three counterfactuals.

Figure D.2: Differential Gains in Reading Test Scores by Achievement and Race Under the Policy Objective of Maximizing Average Reading Achievement

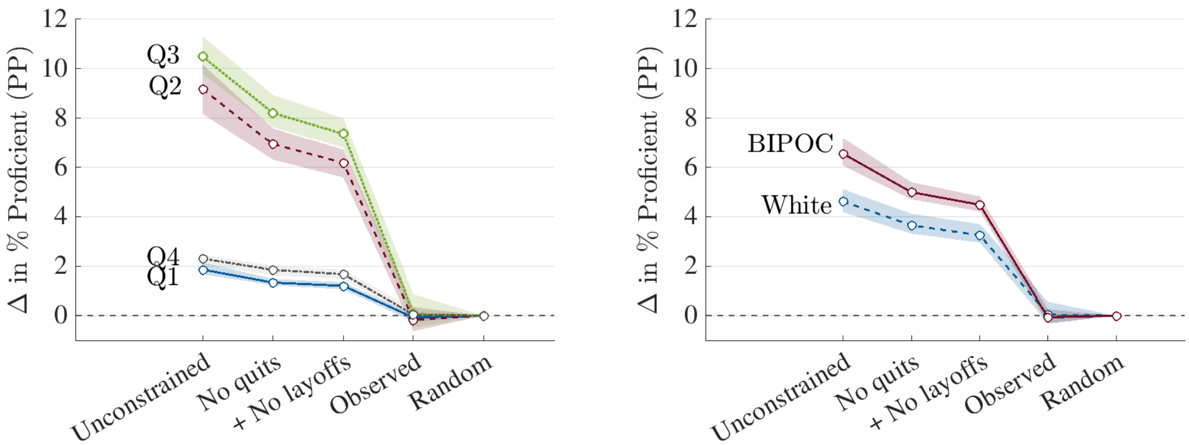


(a) By Achievement Quartile

(b) By Race and Ethnicity

Note: Panels (a) and (b) show the average math test score gains in standard deviation (SD) terms and 95% credible intervals relative to those under the observed and a random assignment by baseline student achievement quartile (Q1-Q4) and student race and ethnicity, respectively, under the three counterfactual scenarios that maximize average math test scores. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The + *No layoffs* counterfactual further restricts the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample.

Figure D.3: Differential Gains in Reading Percentage Proficient by Achievement and Race Under Policy Objective of Maximizing Percentage with Proficient Achievement in Reading



(a) By Achievement Quartile

(b) By Race and Ethnicity

Note: Panels (a) and (b) show the gains in percentage of proficient students in math in percentage points (PP) terms and 95% credible intervals relative to those under the observed and a random assignment by baseline student achievement quartile (Q1-Q4) and student race and ethnicity, respectively, under the three counterfactual scenarios that maximize the percentage of proficient students in math. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The + *No layoffs* counterfactual further restricts the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample.

Table D.1: Counterfactual Summary: Maximize Reading Average Test Scores

	Average Test Score Gains Relative to the Observed Assignment (SD)		
	Unconstrained	No quits	+ No layoffs
All	0.148	0.114	0.101
	[0.13,0.162]	[0.101,0.127]	[0.089,0.112]
By Achievement			
First quartile	0.131	0.099	0.086
	[0.113,0.148]	[0.087,0.114]	[0.075,0.1]
Second quartile	0.133	0.102	0.089
	[0.117,0.149]	[0.091,0.116]	[0.08,0.102]
Third quartile	0.146	0.114	0.1
	[0.127,0.161]	[0.1,0.126]	[0.088,0.112]
Fourth quartile	0.175	0.137	0.123
	[0.152,0.191]	[0.119,0.15]	[0.105,0.136]
By Race			
BIPOC	0.137	0.104	0.091
	[0.121,0.152]	[0.093,0.117]	[0.082,0.103]
White	0.167	0.132	0.118
	[0.145,0.183]	[0.114,0.146]	[0.101,0.131]

Note: This table presents the average reading test score gains relative to the observed assignment in standard deviation terms (SD) for all students as well as for students by baseline achievement (in quartiles) and race in a counterfactual where average student test scores are maximized. 95% confidence intervals in square brackets.

Table D.2: Decomposition of Gains in Reading test score when Maximizing Average Reading Student Test Scores

	Decomposition of Gains Relative to the Observed Assignment (SD)			
	Total	*General*	*Match Effects*	
	Gain	*Effectiveness*	*Observable*	*Unobservable*
	(1)	(2)	(3)	(4)
Unconstrained	0.148	0.014	0.021	0.111
	[0.130,0.162]	[0.009,0.019]	[0.017,0.026]	[0.089,0.126]
No quits	0.114	0.010	0.016	0.087
	[0.101,0.127]	[0.006,0.016]	[0.012,0.019]	[0.070,0.100]
+ *No layoffs*	0.101	0.007	0.014	0.079
	[0.089,0.112]	[0.004,0.011]	[0.011,0.017]	[0.064,0.090]

Note: This table presents the decomposition of the gains in average reading test scores in standard deviation (SD) terms under the three counterfactual scenarios relative to the outcomes under the observed assignment. The *Unconstrained* counterfactual is one in which teachers are assigned to positions to maximize each of the two policy objectives, given the positions and the set of teachers described above, without any additional constraints. The *No quits* counterfactual restricts each teacher's menu of possible assignments to only those that the teacher would weakly prefer over her observed assignment. The + *No layoffs* counterfactual further restricts the teacher candidates for alternative assignments to only those teachers observed to be assigned to a position in our sample. 95% confidence intervals in square brackets.

Table D.3: Counterfactual Summary: Maximize Reading Percent Proficient

	Gains in Percent Proficient in Reading Relative to the Observed Assignment (PP)		
	Unconstrained	No quits	+ No layoffs
All	0.059	0.045	0.041
	[0.053,0.065]	[0.041,0.049]	[0.037,0.044]
By Achievement			
First quartile	0.019	0.014	0.013
	[0.017,0.023]	[0.013,0.016]	[0.011,0.014]
Second quartile	0.093	0.071	0.063
	[0.083,0.108]	[0.064,0.079]	[0.057,0.071]
Third quartile	0.104	0.081	0.073
	[0.093,0.112]	[0.073,0.088]	[0.066,0.079]
Fourth quartile	0.023	0.018	0.016
	[0.021,0.024]	[0.017,0.02]	[0.015,0.018]
By Race			
BIPOC	0.066	0.051	0.046
	[0.059,0.075]	[0.045,0.055]	[0.041,0.05]
White	0.046	0.036	0.032
	[0.042,0.048]	[0.033,0.039]	[0.029,0.035]

Note: This table presents the change in the percent of proficient students relative to the observed assignment in percentage points (PP) for all students as well as for students by baseline achievement (in quartiles) and race in a counterfactual where the share of proficient students is maximized. 95% confidence intervals in square brackets.