NBER WORKING PAPER SERIES

TEACHER LABOR MARKET EQUILIBRIUM AND THE DISTRIBUTION OF STUDENT ACHIEVEMENT

Michael D. Bates
Michael Dinerstein
Andrew C. Johnston
Isaac Sorkin

Working Paper 29728
http://www.nber.org/papers/w29728

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2022, Revised June 2022

Previously circulated as "Teacher Labor Market Equilibrium and Student Achievement." We thank Peter Arcidiacono, Natalie Bau, Christina Brown, Jediphi Cabal, Dennis Epple, Felipe Goncalves, Caroline Hoxby, Peter Hull, Neale Mahoney, Parag Pathak, Camille Terrier, and participants at the 2022 ASSAs, SOLE 2021, NBER SI: Education, CESiFo Education, ASU Empirical Micro Conference, Berkeley, Boston College, Carnegie Mellon, Chicago, Georgetown, Harvard, Michigan, Penn, Stanford, UCLA, UCSD and the University of Minnesota “Big Data” conference for helpful comments and conversations. Thanks to Ian Calaway for research assistance. We also thank representatives at an unnamed district, Kara Bonneau, and the North Carolina Education Research Data Center for data merging and access. Bates thanks the Blum Initiative and Sorkin thanks the Alfred P. Sloan Foundation for support. This research was approved by the Stanford IRB (#43865), the UC Merced IRB (UCM2019-88), and the UC Riverside IRB (HS-16-094). It was deemed non-human subjects research by the University of Chicago IRB (IRB17-1651). Mistakes are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Teacher Labor Market Equilibrium and the Distribution of Student Achievement
Michael D. Bates, Michael Dinerstein, Andrew C. Johnston, and Isaac Sorkin
NBER Working Paper No. 29728
February 2022, Revised June 2022
JEL No. I28,J08,J45

ABSTRACT

We study the equity and efficiency consequences of the allocation of teachers to schools. Within a district, wages are uniform across potential assignments. Because teachers tend to prefer schools with more advantaged students, this uniformity may lead to inequity among students. Because there might be match effects in teaching, this uniformity may lead to inefficient allocations. While we do observe inefficient allocations (there are meaningful gains from reallocation), surprisingly, we do not observe inequity: advantaged and disadvantaged students have teachers with similar value-added. To understand why uniform wages lead to inefficiency but not inequity, we use rich data from the teacher transfer system linked to test score data to estimate an equilibrium model of the teacher labor market. We find that the allocation is equitable because principals hire noisily, tending not to select their most effective applicants. Achieving most efficiency gains, however, requires differentiated wages that compensate teachers for match output.

Michael D. Bates
3135 Sproul Hall
Department of Economics
University of California, Riverside
Riverside, CA 92521
United States
michael.bates@ucr.edu

Michael Dinerstein
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
mdinerstein@uchicago.edu

Andrew C. Johnston
Department of Economics
University of California, Merced
5200 N Lake Road
Merced, CA 95340
and NBER
acjohnston@ucmerced.edu

Isaac Sorkin
Department of Economics
Stanford University
579 Jane Stanford Way
Stanford, CA 94305
and NBER
sorkin@stanford.edu

A data appendix is available at http://www.nber.org/data-appendix/w29728

Teachers matter for student outcomes, and there are large differences in teacher quality even within a single district (e.g., Hanushek, Kain and Rivkin (2004); Rockoff (2004); Chetty, Friedman and Rockoff (2014*a*)). The *allocation* of teachers to schools matters for equity and efficiency. For equity, it affects the achievement gap between advantaged and disadvantaged students. For efficiency, it affects aggregate achievement through match effects and differential class size.

The structure of compensation, however, suggests that the current allocation might depart from equity and efficiency goals. Within a district, wages are uniform across potential assignments (e.g., Biasi (2021)). Because teachers have preferences over school assignments and may choose not to accept an unfavorable assignment (Rothstein, 2015), they may require compensation for differences in amenities. In the absence of compensating differentials, teacher labor supply to a given school is largely a function of amenities like student composition, which may threaten equity as teachers may prefer schools with economically advantaged students. Uniform pay may also leave productive matches unrealized, as teachers and principals are not compensated for output.

In this paper, we study the equity and efficiency of the current within-district allocation of teachers to schools, and what policies would lead to better allocations. We find that our focal district's current allocation is surprisingly equitable—advantaged and disadvantaged students have equally strong teachers—but that there are meaningful aggregate achievement gains from further reallocation. To understand why wage uniformity leads to inefficiency but not inequity, we investigate several channels and find that the allocation is equitable because principals hire noisily, tending not to select their most effective applicants. Noisy principal hiring, though, does not lead to the most productive teacher-school matches, which require differentiated wages that compensate teachers for match output.

We begin by emphasizing a basic puzzle about the current allocation. Consistent with a long literature (e.g., Antos and Rosen (1975)), simple cuts of the data suggest that on average teachers prefer schools with advantaged students. Typically, we would expect that if an employer faces higher labor supply then either it would pay lower wages or else hire higher quality workers. Since a teacher's wages are uniform within a district, we would expect schools with advantaged students to hire better teachers. Instead, and consistent with a recent literature (e.g., Mansfield (2015), Angrist et al. (2021)), we find that the current allocation of teachers to students is surprisingly equitable: disadvantaged and advantaged students have teachers that are equally strong as measured by value-added.[1]

We consider four types of explanations for the surprising parity of the allocation. First, there might be something favorable about how the market clears. Second, there might be preference heterogeneity among teachers such that the marginal teachers have very different preferences than

[1]This literature is slightly mixed, which hints that market institutions matter. See Isenberg et al. (2022), Chetty, Friedman and Rockoff (2014*b*), and Sass et al. (2012) for papers finding no or small gaps. See, e.g., Goldhaber, Lavery and Theobald (2015) and Goldhaber, Theobald and Fumia (2022) for papers using Washington state data that find larger gaps. Angrist et al. (2021) find similar school value-added for advantaged and disadvantaged students.

the average teachers. Third, it might be that principals do not take advantage of the extra supply to hire better teachers. Finally, teachers might have significant comparative advantage with certain students such that schools may be able to find similarly effective teachers.

To evaluate these explanations, we use rich data from the teacher transfer system in a single district linked to test score and class assignment data in North Carolina to estimate an equilibrium model of the teacher labor market. The data allow us to estimate each of the key inputs to the model—market clearing, teacher preferences, principal hiring behavior, and match effects—with relatively weak assumptions. The model allows us to quantify how these factors translate into equilibrium allocations.

In our model, teachers and schools meet and form matches. Teachers can only match with school openings that are active at the same time. Teachers have non-wage preferences over positions, and principals serve as hiring intermediaries who rank teachers on behalf of the district. Each match generates student achievement based on the teacher's value-added at the school. To predict the equilibrium matches, we use the concept of pair-wise stability (Roth and Sotomayor, 1992; Hitsch, Hortaçsu and Ariely, 2010; Banerjee et al., 2013; Boyd et al., 2013).[2]

We specify a multi-dimensional value-added model where teachers may have absolute advantage and comparative advantage in teaching specific student types (Condie, Lefgren and Sims, 2014; Delgado, 2021; Bau, Forthcoming; Biasi, Fu and Stromme, 2021). We divide students based on whether they are economically disadvantaged.

To identify teachers' non-wage preferences over positions (Barbieri, Rossetti and Sestito, 2011; Engel, Jacob and Curran, 2014; Bonhomme, Jolivet and Leuven, 2016; Fox, 2016; Johnston, 2021), we argue based on institutional features and analysis of application behavior that teachers apply non-strategically to positions they prefer relative to their outside option. We specify a rich characteristics-based model of teacher utility with observed and unobserved preference heterogeneity. Teachers on average prefer schools with more advantaged students, though we also find significant cross-teacher heterogeneity.

We then estimate principals' valuations of teachers (Ballou, 1996; Boyd et al., 2011; Jacob et al., 2018; Jatusripitak, 2018; Hinrichs, 2021). Our data include principals' ratings of candidates, which allows us to infer valuations with minimal assumptions despite the possibility that principals act strategically. Using the observed set of applications to each position, we model whether principals give an application a positive rating.[3] Principals value teachers based on observable characteristics like having a graduate degree that poorly predict value-added. From the econometrician's perspective, ratings and value-added are largely independent.

Finally, we estimate market timing based on applicants' and vacancies' periods of activity in

[2]Our model fits in a recent literature considering allocation problems with non-choice outcomes (Agarwal, Hodgson and Somaini, 2020; Ba et al., 2021; Cowgill et al., 2021; Dahlstrand, 2021).

[3]Principals might interview or offer strategically by passing on preferred teachers who are unlikely to accept an offer. As long as these teachers receive at least a positive rating, we can model positive outcomes as non-strategic choices.

our administrative records.

Among the explanations we consider, only principal behavior explains the surprisingly equitable current allocation. While we find some match effects, there is minimal sorting in the current equilibrium. While we find substantial heterogeneity in teacher preferences, it does not help achieve an equitable allocation. In terms of market clearing institutions, we find that there is essentially a unique stable equilibrium, so there is not favorable equilibrium selection. We also find that alleviating timing constraints is not favorable for disadvantaged students. In terms of principals, even though principals place some weight on value-added, its ability to explain the rating decision is limited. This feature of behavior "pushes back" on teacher preferences to generate parity. To quantify this point, we show that if principals were to only place weight on output, then the allocation would be as inequitable as we might expect based on average teacher preferences. Thus, principal behavior explains the surprisingly equitable allocation.

While the current allocation is equitable, policy-makers may prefer allocations that yield higher total achievement or that further close baseline achievement gaps. We find that there are large potential efficiency gains, 0.05 of a student standard deviation (σ), from reallocating teachers across schools. This first-best allocation exploits two sources of efficiency: sorting on teachers' comparative advantage with dis/advantaged students and placement of better teachers in larger classes. Just under half of these gains are from sorting teachers based on comparative advantage (Delgado, 2021). These gains are equivalent to 39% of a standard deviation in teacher value-added or an additional year of experience for a novice teacher, and are larger than the 0.012σ gains we estimate from replacing the bottom 5% of teachers with the median teacher (Staiger and Rockoff, 2010; Neal, 2011).

If the district only valued closing baseline achievement gaps, then it could close about a seventh of the gap between advantaged and disadvantaged students each year.

We use our equilibrium model to evaluate the effectiveness of commonly proposed policies. As before, we find little role for equilibrium selection since there is nearly always a unique stable equilibrium. Complete market coordination such that teachers can apply to any position in a cycle—not just those concurrently active—has a somewhat larger effect, moving student achievement 15% of the way from the status quo toward the first-best.

Policies that target principal behavior by, e.g., giving them information and/or incentives to only rank teachers based on value-added, would have negative equity and efficiency consequences. This result carries over the logic from principals' effect on the current allocation: current principal behavior is a second-best solution (Lipsey and Lancaster (1956)) to the problematic structure of uniform wages amidst strong teacher preferences. Policies that align principal ratings with value-added lead to more homogeneous rankings of teachers so that the highest absolute advantage teachers have many options from which to choose. When teachers rank schools according to their estimated preferences, in which output plays a small role relative to a school's student body composition, increased choice leads to misallocation.

These findings leave the class of policies that varies teacher compensation with the assignment. We compare policies tying teacher compensation to the value-added produced with policies that tie teacher compensation to the fraction of disadvantaged students they teach. When principals hire as in the status quo—i.e., noisily—compensation tied to value-added performs better for both equity and efficiency goals. This occurs because expansions of labor supply to schools with disadvantaged students only closes achievement gaps if principals hire the best teachers, while expansions of labor supply among only the *high quality* teachers changes the composition of labor supply. In contrast, if principals hire based on value-added (with an information and incentives intervention), then compensation tied to the fraction of disadvantaged students outperforms value-added compensation on both efficiency and closing achievement gaps. Thus, changes in teacher compensation are the primary way to move significantly toward the first-best allocations, but the optimal form of compensation depends on principal behavior.

The final policy lesson is that there is a limit to how much of the efficiency gains can be achieved by simply rewarding teachers and/or principals for value-added. The reason is that in the first-best teachers are allocated based on comparative advantage, whereas value-added compensation mean that principals rank in part on absolute advantage.[4] Our estimates show that maximal value-added compensation can only achieve about three-quarters of the efficiency gains, and the remaining quarter requires flexible prices.

Literature: We are not the first paper to point out that principals do not perfectly select teachers to maximize value-added of their students (Ballou (1996) might be the first). What is novel is embedding this observation in an equilibrium framework and showing its consequences for the allocation of teachers to schools.

This paper fits in an emerging literature that uses teacher labor market equilibrium models to assess the gains in student achievement from various policies. These papers range in the allocation problem they consider from national (Combe, Tercieux and Terrier, Forthcoming; Bobba et al., 2021; Combe et al., 2021) to state cross-district (Biasi, Fu and Stromme, 2021) to local within-district (Boyd et al., 2013; Bates, 2020; Laverde et al., 2021) to sectoral (Tincani, 2021). Bau (Forthcoming) studies an equilibrium model of school competition with school-student match effects.

Relative to this literature, our unique combination of detailed data on teacher applications, principal ratings, and student-teacher classroom assignments allows us to identify two-sided heteroge-

[4] To see why the stable and first best allocations can be different, suppose that teacher 1 has output $\{10,9\}$ at schools 1 and 2, respectively, and teacher 2 has output $\{8,0\}$ at schools 1 and 2. Then in any stable equilibrium where both teachers and principals only value output, teacher 1 is assigned to school 1 and teacher 2 is assigned to school 2. In contrast, in the first best, teacher 1 is assigned to school 2 and teacher 2 to school 1. This assignment reflects teachers' comparative advantage. If the comparative advantage of teacher 2 is strong enough, say, her output is $\{11,0\}$, then the decentralized and first-best allocations coincide.

neous preferences and a multi-dimensional production model with straightforward assumptions on behavior. Such data are necessary for our conclusions. To determine that the explanation for the current allocation's equity is that principals do not select teachers solely based on value-added, we need to observe the applications that the principal receives as well as choices the principal makes. In contrast, Boyd et al. (2013), Biasi, Fu and Stromme (2021), and Tincani (2021) do not have data on applications and so assume that principals (school districts) hire the highest value-added teachers subject to a budget constraint. We also note that when principals do rely on teacher observable characteristics for hiring, these poorly predict value-added. In contrast, Combe, Tercieux and Terrier (Forthcoming), (Tincani, 2021) and Bobba et al. (2021) lack student-teacher assignment data and use teacher observable characteristics to infer teacher output. Further, our teacher application data allow us to estimate significant preference heterogeneity and the joint distribution of teacher preferences and value-added. In contrast, Boyd et al. (2013), Biasi, Fu and Stromme (2021), Bates (2020), and (Tincani, 2021) do not have such data and infer preferences based on equilibrium allocations.

While the model is tailored to our empirical setting, the case study carries important lessons for broader analysis of labor markets. First, related to the literature on non-pay amenities (e.g., Rosen (1986), and Sorkin (2018)), the fact that teachers express strong preferences over positions in the absence of wage variation is clear evidence of the important role of amenities in shaping worker choices. Second, as Oyer and Schaefer (2011) emphasize, labor economists should study how firms hire workers. In this paper, we estimate an empirical model of the hiring decision and find that the "mistakes" in hiring have desirable consequences for the overall allocation. Third, as Card et al. (2018) emphasize, the empirical labor literature on imperfect competition would benefit from "IO-style" case studies of particular markets. Here, we have a context where we are able to estimate preferences of both sides of the market with rich data and link preferences directly to output.

1 An equilibrium model of the teacher labor market

In this section, we describe an equilibrium model of the teacher labor market within a school district. The model clarifies the set of factors shaping the equilibrium, and thus the set of potential explanations for the parity of the current allocation. Having specified the model, we then use it to study counterfactuals.

We defer a full description of the empirical setting to Section 4, but we highlight a few features that inform the model. First, the teacher labor market operates on a rolling basis from April to August each year. This segments the market in time. Second, the market is decentralized such that teachers choose which positions to apply to, and principals choose whom to interview and then whom to offer jobs. Finally, a teacher's wage does not vary based on the offer she accepts.

1.1 Set-up

We begin with specialized notation that we generalize in the next section. Denote teachers by j, and schools and principals by k. Teachers and principals both receive quasi-linear utility from an assignment. Teacher j derives utility u_{jk} from teaching at school k:

$$u_{jk} = \tilde{u}_{jk} + w_{jk}, \tag{1}$$

where w_{jk} is the wage, and $\tilde{u}_{jk}$ is a match-specific amenity. School k (or the principal who runs it) derives utility, v_{jk}, from hiring teacher j. This utility is linear in a non-wage component less the wage paid to teacher j:

$$v_{jk} = \tilde{v}_{jk} - w_{jk}. \tag{2}$$

A teacher-school assignment produces student value-added VA_{jk}. Below, we specify a functional form for VA_{jk}.

Finally, let $\mathcal{J}$ be the set of teachers, $\mathcal{K}$ be the set of schools, and assume for simplicity that the number of teachers and schools is the same. An assignment of teachers to classrooms is a one-to-one and onto function (bijection): $\phi : \mathcal{J} \to \mathcal{K}$ so that $\phi(j) = k$, the school k to which teacher j is assigned. Denote by Φ the set of all possible assignments.

1.2 Decentralized equilibrium

To characterize the current equilibrium, we use a decentralized equilibrium concept derived from the literature on matching. Schools meet with all teachers who are in the market at the same time. The equilibrium concept is pair-wise stability. Under a stable allocation, no teacher and school pair would prefer to jointly deviate and match (Roth and Sotomayor (1992), Definition 2.3).

To model the empirical status quo, we assume (1) teachers and principals have the preferences we estimate for them and (2) the timing of the market follows that which we observed in the administrative records, where not all matches are feasible. There is not necessarily a unique stable equilibrium. We model the status quo using the teacher-proposing deferred-acceptance algorithm (DA). For reasons that we explain in the next section, we use DA in order to find stable equilibria, not because DA is actually used in this market.

1.3 Explanations for the current allocation

We are interested in understanding the parity of the current allocation, despite the average teacher preferring schools with advantaged students (a fact we will establish in Sections 3 and 5). The model embeds five potential explanations.

The first explanation is that there are multiple possible equilibria and the current equilibrium is favorable to disadvantaged students. We can quantify the possibility of this explanation by examining the range of stable equilibria. The teacher-proposing DA is the teacher-optimal stable allocation (Roth and Sotomayor (1992), Corollary 2.14) meaning every teacher would (weakly) prefer their assignment to that in every other stable allocation. In contrast, the school-proposing DA (where schools only value output) is the best stable allocation in terms of student achievement. By examining these extremal equilibria, we are able to consider whether there are alternative equilibria with different distributional consequences.

The second explanation is that the timing of when teachers and positions are active is favorable to disadvantaged students. For example, positions with many disadvantaged students might be active in a moment when only good teachers are active. As a way to quantify this possibility, we ask what the allocation would look like if all positions and all teachers were active at the same time.

The third explanation is that average teacher preferences hide substantial heterogeneity. For example, there is a subset of high quality teachers with strong preferences to teach at schools with disadvantaged students. We allow for this possibility by estimating preference models allowing for rich forms of heterogeneity across teachers that is potentially correlated with value-added. We can then ask whether the allocation is different if we instead impose simpler teacher preference models.

The fourth explanation is that in generally attractive schools, principal hiring behavior does not take advantage of the excess demand to hire better teachers. This behavior could reflect some combination of lack of information (principals do not know who the high value-added teachers are) and incentives (principals recognize the high value-added teachers, but have other priorities and so choose not to hire them). We allow for this possibility by comparing equilibria where principals behave according to our estimates, and those where they maximize value-added.

The fifth explanation is that teachers have significant comparative advantage with different student types. If teachers who are effective with advantaged students are not effective with disadvantaged students, and vice versa, then all schools may be able to find teachers well-suited for teaching their students. We incorporate this possibility by specifying production models with and without match effects and comparing the predicted achievement in the respective status quo equilibria.

1.4 First-best problems

Beyond describing the properties of the current equilibrium, we are interested in policies that increase aggregate achievement. Thus, as a benchmark, we consider a school district's first-best assignment problem. Here we take as given the set of teachers and positions the district has and ask how to assign them. In Section 3, we consider just the teachers in the district. In Section 7, we consider the set of teachers who apply in the transfer system and for whom we can estimate value-added: this set includes teachers who have previously taught anywhere in the state. If we considered all possible teachers in the single district's problem (including potential teachers and those who do not apply to the district), then we would be ignoring how our focal district's behavior affects the allocation of teachers to and within other districts. The allocation problem then would no longer map into a social planner's problem.

The district values students' outcomes and teachers' preferences over assignment (non-wage utility):

$$\max_{\phi \in \Phi} \{\omega \sum_{j \in \mathcal{J}} VA_{j\phi(j)} + \sum_{j \in \mathcal{J}} \tilde{u}_{j\phi(j)}\}. \tag{3}$$

To understand this allocation problem, note that the the first term ($\sum_{j \in \mathcal{J}} VA_{j\phi(j)}$) is the output achieved given an assignment ϕ. Here, the district weights the output of all students equally; below we consider extensions where the district places different weight on different types of students (e.g., disadvantaged and advantaged students). The second term ($\sum_{j \in \mathcal{J}} \tilde{u}_{j\phi(j)}$) is the total amenity value that teachers gain from this allocation. Finally, ω is the weight that the district places on student achievement relative to teacher preferences. We will often evaluate allocations solely in terms of student achievement ($\omega = \infty$).

We exclude principal behavior from the district's value of an allocation to focus on the essential elements of the problem. Specifically, the district could plausibly bypass the intermediary of the principal and directly hire for schools. In this sense, we do not commit to a utility interpretation of principals' preferences, and could instead interpret them simply as a hiring rule; hence, we tend to use the language of principal behavior.

Because the paper's goal is to study the allocation of teachers, and not how best to use existing dollars, we do not include a budget constraint in the district's problem. As cost is still a relevant consideration in evaluating allocations, in Section 8 we compare the effectiveness of policies that cost equal amounts.

We consider a range of district first-best problems where the relative weight on students varies. We refer to the resulting set of optimal allocations as the production possibilities frontier. The slope of the frontier captures the trade-off between student achievement and teacher utility.

1.5 Policies

We consider five policies that a district might pursue to attempt to reach a first-best allocation.

The first four policies parallel the first four factors generating equilibrium described above. Notably, versions of these policies have all been proposed or implemented in different districts in the United States. First, the district might wish to affect equilibrium selection.[5] Second, the district might try to change the timing of the market.[6] Third, the district might provide incentives and/or

[5] Some districts have focused on changing the algorithm that clears the market (Davis, 2021).

[6] Some districts allow some schools to hire first Kraft et al. (2020), and others advocate for shifting the timing to

information to teachers to make them value schools based on the teachers' potential output at the school.[7] Fourth, the district might provide incentives and/or information to principals to make them hire teachers on the basis of value-added.[8]

Finally, all four of these policies—even in combination—are not necessarily sufficient to achieve the first-best allocations described in the previous section. The reason is that even when teachers and principals only value output, the allocation sorts teachers based in part on absolute rather than comparative advantage. To achieve the first-best allocations, it is sufficient to have the following combination of policies: (1) districts inform principals and compensate them for output (so that principals rank teachers by match-specific value-added, VA_{jk}), (2) all teachers and schools are in the market simultaneously, and (3) wages may vary with each teacher-school pair. This last policy makes utility transferable. Whereas output bonuses only let wages vary depending on the output in the assigned position, flexible wages would let wages depend on a teacher's output in other assignments, teacher preferences, and the distribution of other teachers' potential output and preferences. With full information, having this flexibility guarantees that the district can implement any first-best allocation (Shapley and Shubik, 1971).

Other than for equilibrium selection, there is no theorem that the other policies in isolation necessarily improve output. The theory of the second-best states that when we are away from the first-best allocation because of multiple factors, then fixing any one factor can worsen outcomes.

1.6 Empirical plan

The model highlights the empirical objects we need to estimate to be able to simulate the impact of the above policies and understand the current allocation. We start by estimating the potential outcomes of teachers across schools, VA_{jk}. We then estimate teachers' amenity value across all assignments, $\tilde{u}_{jk}$, and principals' non-wage utility from hiring each teacher, $\tilde{v}_{jk}$. Finally, we model which positions were available to each teacher in the observed equilibrium.

2 Data

To estimate the objects of interest, we use rich data on the labor market for elementary school teachers. The first type of data comes from the platform used to hire teachers in our focal district. We use this data to estimate teacher and principal preferences. The second type of data comes

benefit certain types of schools (Levin and Quinn, 2003).

[7]Examples of teacher-level output bonuses include Indiana (Marcotte, 2015) and the ProComp policy in Denver (Atteberry and LaCour, 2020). North Carolina implemented bonuses for teaching in hard-to-staff schools from 2001-2004 (Clotfelter et al., 2008) while South Carolina provides high poverty districts with funding for teacher bonuses (Fox, 2017).

[8]Examples of principal-level bonuses or information treatments include North Carolina, which instituted principal bonuses as a function of the growth in student test scores in 2017-2018 (Pridemore, 2017), and New York City, which piloted a program giving principals information about their teachers' performance in 2007-2008 (Rockoff et al., 2012).

from staffing and achievement records from state accountability records. This data provides us with student-level test score data that we link to teachers and use to estimate value-added models. In addition, these records provide information about a variety of demographic characteristics of teachers and students as well as teachers' education and experience in the district. In this section, we briefly describe the data. See Appendix A for further details and Appendix Table A1 for summary statistics across samples.

2.1 Job application and vacancy data

We obtained application records from our focal district's system, which spans 2010 through 2019 and records 346,663 job applications. In the system, schools post job vacancies, and applicants apply for jobs. The system also records various actions that principals take.

For every posted position, the vacancy files indicate the school, position title, and whether the position is full-time or part-time. We use the detail on the position title to isolate non-specialized elementary school teacher jobs (i.e., we omit elementary school jobs such as "literary facilitator elementary").

We use two features of the teacher file. First, the file records which vacancies the candidate applied to, and when she submitted the application. The timing information allows us to construct choice sets, which we detail in Section 4. Second, the file records the city, zip code, and, in some cases, exact address where the teacher lives. This feature allows us to construct the commute time for each teacher-position combination.

We also have data in which principals record their assessments of teachers. Principals record their interest in different applicants, the equivalent of a "good" and a "bad" pile. Principals also often record which candidates they invited to interview, which candidates were offered the position, and which candidates were hired.

2.2 Administrative data

We link the platform data to state administrative records on teachers and students. For teachers, we have their experience, salary, licensing, certification scores, class assignments, and the school where they work. For students, we have scores on standardized exams, grades, race, sex, and whether they qualify as disadvantaged based on Federal programs. Records on class assignments allow us to link teachers to students.

The North Carolina Education Research Data Center (NCERDC) matched the data from the job-market platforms to the state's administrative data. They performed an interactive fuzzy match using names and birth year. For teachers who had a sufficiently good match (that is, a unique name-birth-year combination), we have a de-identified ID that allows us to connect their platform data to their staffing records and students' achievement.

3 Production of student achievement

In this section, we first specify the production model, which specifies teacher output at each school. Second, we describe our three-step estimation procedure and discuss parameter estimates. Third, we present a range of validation checks. Fourth, we describe important properties of the current allocation. Finally, we use our estimates to show that reallocating teachers across schools can produce meaningful gains in aggregate achievement and close baseline achievement gaps.

3.1 Model

We want to predict how teacher output would change depending on the teacher-school match. While the teacher value-added literature usually estimates constant effects models, we follow the quickly expanding literature documenting match effects and allow for comparative advantage (Dee, 2004, 2005; Jackson, 2013; Aucejo et al., 2021; Delgado, 2021; Graham et al., 2020; Biasi, Fu and Stromme, 2021). This serves two purposes. First, substantial match effects could explain the parity of the current allocation. Second, match effects would likely increase potential achievement gains from reallocation.

We specify a model of match effects that is identified with our data and allows us to make output predictions in unobserved matches. Since a teacher typically works in just a few schools during her career, we cannot identify fully flexible match effects. Instead, we use low-dimensional match effects where teachers have different value-added with students of different observable types; here, we focus on a single student characteristic—economic disadvantage. Thus, a teacher's school-level match effect depends on the observed demographic composition of the school and the teacher's comparative advantage with each type of student.

We use notation that follows Chetty, Friedman and Rockoff (2014*a*) and Delgado (2021). Let i index students and t index years, where t refers to the spring of the academic year, e.g., 2016 refers to 2015-2016. Each student i has an exogenous type $m(i,t) \in \{0,1\}$ in year t (whether the student is economically disadvantaged). Student i attends school $k = k(i,t)$ in year t and is assigned to classroom $c = c(i,t)$. Each classroom has a single teacher $j = j(c(i,t))$, though teachers may have multiple classrooms.

Student achievement depends on observed student characteristics, teacher value-added, school effects, time effects, classroom-student-type effects, and an error term. Formally, we model student achievement A^*_{it} as:

$$A^*_{it} = \beta_s X_{it} + \nu_{it} \tag{4}$$

where X_{it} is a set of observed determinants of student achievement and

$$\nu_{it} = f(Z_{jt}; \alpha) + \mu_{jmt} + \mu_k + \mu_t + \theta_{cmt} + \tilde{\varepsilon}_{it}. \tag{5}$$

Here, Z_{jt} is teacher experience (and f maps experience into output) and μ_{jmt} is teacher j's value-added in year t for student type m, excluding the return to experience. As in Chetty, Friedman and Rockoff (2014*a*), we allow a teacher's effectiveness to "drift" over time. μ_k captures school factors, such as an enthusiastic principal, while μ_t are time shocks. θ_{cmt} are classroom shocks specific to a student type, and $\tilde{\varepsilon}_{it}$ is idiosyncratic student-level variation.

We make three assumptions, which are standard in the literature (see Appendix B for formal statements of these assumptions). The first assumption is that classroom-student type shocks and idiosyncratic student-level variation are orthogonal to teacher and school assignments and follow a stationary process. We allow classroom shocks to be correlated across student types in the same classroom, but restrict all cross-classroom or cross-year correlations in shocks to be zero.

The second assumption is that the non-experience part of teacher value-added for each student type follows a stationary process that does not depend on the teacher's school. We also assume that the covariances between the teacher's value-added across student types depend only on the number of years elapsed.

The third assumption is that drift and school effects are independent. This assumption rules out teacher mobility (or initial assignments) related to the drift of the teacher's effect. We still permit teacher-school assignments to be non-random, and quite possibly related to a teacher's comparative advantage in teaching different student types.

Our object of interest is a forecast of teacher j's value-added from a hypothetical assignment to a new classroom (or set of classrooms) in school k. Define p_{kmt} as the proportion of type-m students in school k in year t. Given our low-dimensional model of match effects, a teacher's predicted mean value-added at school k in year t is:

$$VA^{p}_{jkt} = p_{k0t}\mu_{j0t} + p_{k1t}\mu_{j1t} + f(Z_{jt};\alpha), \tag{6}$$

such that a teacher's total value-added for n_{jkt} students is $VA_{jkt} = n_{jkt}VA^{p}_{jkt}$. We use data through $t-1$ from the whole state to forecast VA^{p}_{jkt} for assignments we see in the data and for counterfactual assignments. For the observed assignments, we forecast the teacher's value-added were she to receive a new draw of students and classrooms at that school. For the counterfactual assignments, we predict a teacher's value-added for schools at which she did not teach.

3.2 Estimation

We estimate our model in three steps using math scores.[9] In the first step, we estimate the coefficient on characteristics by regressing test scores (standardized at the state-level to have mean 0 and standard deviation 1 in each grade-year) on a set of student characteristics and classroom-student-type

[9]Focusing on a single subject allows us to rank all possible levels of output. We follow Biasi, Fu and Stromme (2021) in choosing math because it is typically more responsive to treatment (e.g., Rivkin, Hanushek and Kain (2005), Kane and Staiger (2008), and Chetty, Friedman and Rockoff (2014*a*) for evidence).

fixed effects. In the second step, we project the residuals (A_{it}) onto teacher fixed effects, school fixed effects, year fixed effects, and the teacher experience return function. In the final step, we form our estimate of teacher j's value-added in year t for type m (μ_{jmt}) as the best linear predictor based on the prior data in our sample (this prediction includes the experience function). Since in this final step we shrink the estimates, we understate the dispersion in match effects relative to the true dispersion. That said, using shrunken estimates and prior data means that we use the information available to policy-makers. See Appendix B.2 for estimation details and a discussion of what variation pins down parameters.

Alternative value-added models: We consider three alternative value-added models. The first is a homogeneous effects model, where we assume that teachers' effects on students and the classroom shocks are type invariant. The second model estimates the school effects differently: rather than including school fixed effects (as in, e.g., Jackson (2018)), we include school-level means of all of the covariates (as in, e.g., Chetty, Friedman and Rockoff (2014*b*). Third, we include teacher-year fixed effects in the residualization step, rather than teacher-class-student type effects as in our baseline. See Appendix B.3 for details.

3.3 Validation of the match effects model

To test whether our estimates of teacher comparative advantage with different types of students simply reflect statistical noise, we perform three tests of our multi-dimensional value-added model versus a single-dimensional model. First, we estimate confidence intervals for the structural parameters in our production model. Second, we perform a likelihood-ratio test comparing our model to a model with one-dimensional teacher value-added. Third, we test whether teachers who have previously been stronger with disadvantaged students see increases in estimated value-added when transferring to schools with greater shares of disadvantaged students. Similarly, we test the reverse relationship. If our comparative advantage estimates only reflected spurious relationships, then they would not predict changes in output upon transfer. All three of these tests allow us to reject homogeneity. See Appendix B.4 for further details on all three tests.

To validate our value-added model, we slightly modify Chetty, Friedman and Rockoff (2014*a*)'s test for mean forecast unbiasedness. We predict a teacher j's value-added in school k in year t (μ_{jkt}) using data from all years prior to t. We then regress the realized mean student residuals in year t ($\bar{A}_{jt}$) and test whether the coefficient on our prediction equals 1. Column (1) of Table 1 shows that the math value-added estimate is an unbiased predictor of residualized output, with a tight confidence interval around 1.05. Figure 1 shows that the forecast unbiasedness holds throughout the distribution of teacher value-added.

We conduct a similar test for the comparative advantage component of value-added, which will be important for explaining the current allocation's parity and potential efficiency and distributional reallocation gains. If teachers' heterogeneous effects by student type vary with the environment—for instance, teachers might target instruction toward the median student in the class—then our model may poorly forecast a teacher's comparative advantage. In column (2) we compare our forecast of the difference in a teacher's value-added across (economically) disadvantaged and advantaged students with the realized test score difference. Again, we find that our estimates are nearly forecast unbiased. Appendix Figure A1 shows that the forecast unbiasedness holds throughout the distribution.

We perform three tests of whether our measure of teacher value-added also forecasts output across the types of teacher moves that we consider in counterfactuals. Our motivation in specifying a low-dimensional model of match effects is that we do not observe a teacher's potential outcomes at all schools, and so we cannot directly assess the quality of our model across all potential outcomes. What we can do, however, is look at the types of changes in the data that are closest to those that we contemplate in counterfactuals. First, we consider moving teachers across schools. Second, we consider moving teachers across classrooms (schools) with large changes in classroom composition in terms of advantaged and disadvantaged students. Third, we consider moving teachers across classrooms (schools) with different numbers of students.

In the spirit of Chetty, Friedman and Rockoff (2014*a*)'s quasi-experiment, how well do our measures predict output when teachers switch schools? In column (3) of Table 1 we find no systematic change in value added after transferring schools.[10] We also test how well our value-added estimates predict transfer effects in column (4). We find that our value-added measure is mean forecast unbiased.

How well do our measures predict output when there are large changes in student composition? We split the data into three groups based on the size of the change between the estimation sample (before year t) and the prediction sample (year t) in the share of disadvantaged students. To examine the validity of our prediction in extreme reassignments, we look at changes below the 10th percentile, above the 90th percentile, and between the 10th to the 90th percentiles. For large negative changes (in Column (5) of Table 1), we find that our measure is forecast unbiased while we find a small forecast bias for large positive changes.

How well do our measures predict output when there are large changes in class size? We perform a parallel analysis for class size as we did for student composition and find similar answers. Specifically, in Column (6) of Table 1 we find that for large negative changes our measure is forecast unbiased, while for large positive changes we find slight evidence of forecast bias.[11]

[10] In the same setting but in an earlier sample, Jackson (2013) estimates increases in output following transfers. In our later sample, we see zero or negative effects when estimating Jackson (2013)'s event study specification (Appendix Figure A2).

[11] The regressions assess forecast unbiasedness with a linear model. We show how our predictions perform nonparametrically in Appendix Figures A3a and A3b for large decreases and increases in class size, respectively. We see that our predictions are forecast unbiased throughout the quality distribution such that quality differences across teachers are likely to remain across schools with different class sizes. Note that we are not ruling out class size effects, but rather that

Our model's ability to predict value-added across settings with minimal bias instills confidence that we can predict the production effects from counterfactual allocations of teachers to schools.

3.4 Properties of the current allocation of teachers to schools

We highlight several features of the current allocation of teachers to schools in our focal district. We also show that these patterns are similar in other districts in North Carolina and, for the subset of the patterns that we can compute, that these are similar in a national dataset.

Table 2 suggests that on average teachers do not like teaching at schools with disadvantaged students. Consistent with teachers moving away from such students over time, disadvantaged students have less experienced teachers. More directly, attrition both from the school and the district is higher for teachers of disadvantaged students. This finding holds outside our focal district.

Typically, if an employer is more desirable (i.e., schools with more advantaged students), then—in the absence of varying prices—we would expect it to have better workers (i.e., teachers with higher value-added). Advantaged schools do employ teachers who look better along many dimensions that a naive observer would expect to predict higher test-score value-added: they are more experienced, score higher on the Praxis exam, more likely to be licensed, have a graduate degree, and to be NBPTS certified. While these observed variables correlate with value-added in the expected ways (and often with statistical significance), Appendix Table A2 shows that their explanatory power is very limited: the R^2 for either an aggregated measure of value-added or a type-specific measure of value-added is below 0.025.

Advantaged students do not have higher value-added teachers than disadvantaged students, which is puzzling given the labor market facts we documented above. In levels, the achievement gap is large: advantaged students score 0.86σ higher than disadvantaged students. In gains, however, we find that disadvantaged students experience identical test score growth. More generally, consistent with identical test score growth, Appendix Figure A4 shows that the level gaps are similar across grades. Given that test score growth is used to estimate teacher value-added, it is perhaps unsurprising that we find almost no difference in the quality of teachers of advantaged and disadvantaged students. This result holds for our baseline value-added measure when we measure quality by ability with either advantaged or disadvantaged students, as well as our alternative value-added measures.

That disadvantaged students do not have worse teachers is not unique to our district and is consistent with others' findings. Columns (3) and (4) of the Table show that we find similar patterns of faster test score gains for disadvantaged students and better teachers outside of our focal district as well. Like us, Sass et al. (2012, Table 2) shows that in both North Carolina and Florida schools with greater share of disadvantaged students experience faster test score gains.[12]

these enter the school fixed effects, and we assume that class size effects do not interact with the teacher's identity.

[12]In Appendix Table A3 we report summary statistics dividing schools by share disadvantaged, which is how Sass

We can now address one of the potential explanations for the current allocation's parity: sorting on comparative advantage. In our estimated production model, a teacher's effectiveness across student types is highly correlated (0.86, Appendix Table A4), though there is still some room for comparative advantage. In Table 2 we show a teacher's type-specific value-added, split by the student type taught. The average disadvantaged student has a teacher who is equivalently effective with both student types (0.02) while the average advantaged student has a teacher who has slight comparative advantage with *disadvantaged* students. Thus, we can rule out sorting on comparative advantage as an explanation for the current allocation's parity.

Figure 2 shows two properties of the current allocation that shape the gains from reallocation and are not unique to our focal district. First, disadvantaged students are in schools with smaller class sizes . The second is a more novel result, comparative advantage with disadvantaged students is positively correlated with absolute advantage within the district and state as a whole. In terms of the optimal allocation, these properties push in opposite directions. The first suggests sending the stronger teachers to the advantaged students (in larger classes). The second suggests sending such teachers to the disadvantaged students (where they have comparative advantage). In Appendix Figure A5, we show that the fact that disadvantaged students are in smaller classes within each district is true nationwide (because we use different data for this exercise, we reproduce the statistic for our focal district in this alternative dataset). We also compute the two correlations discussed previously in all the districts in North Carolina. The figure shows that our focal district is towards the center of the distribution across districts along both dimensions.

3.5 Gains from alternate allocations

Efficiency objectives: We solve the district's problem in Equation 3 where the district only values student output. Table 3 shows that there are sizeable output gains from hypothetical re-allocations of all teachers in our district in 2016. To focus solely on gains from reallocating teachers across schools, we give each classroom within a school the same composition and number of students. The top panel shows per-student gains or losses (column 1) from movements to the output-maximizing ("best") or output-minimizing ("worst") allocations. The per-student gain of 0.054σ in the output-maximizing allocation reflects improved sorting of teachers to schools without changing the set of available teachers. The actual allocation is only slightly better than randomly assigning teachers to schools (row 2 of Table 3), and the range of annual output between the best and worst allocations is 0.11σ (row 1 minus row 3 of Table 3).

Are these gains large or small? One way of contextualizing these gains is to compare them with the cross-sectional standard deviation of predicted teacher value added in our district, which is about 0.14σ (see Appendix Figure A6 for the distribution). The effect of the first-best reallocation

et al. (2012, Table 2) is structured. Unlike us, they find that the value-added of teachers in schools with many disadvantaged students is worse, which is slightly puzzling given they document a lack of differences in raw test score growth.

is to increase teacher output by about a third of a standard deviation.

Another way of contextualizing the size of the gains is to compare them to the impacts of two commonly-proposed policies or allocation rules. Reallocations within schools—i.e., matching teachers based on within-school variation in classroom composition—would only achieve 28% of the gains from the cross-school gains (row 4). These within-school gains reflect the fact that the within-school assignment of teachers to classrooms exploits none of these gains and produces the same output as random allocations (row 5). Replacing the bottom 5% of teachers, where we rank teachers based on their forecasted value added in their actual assignment, with teachers of median quality (as in Hanushek (2009) and Chetty, Friedman and Rockoff (2014*a*)), would achieve 22% of the gains from better sorting of the existing teacher pool (row 5).

In our reallocation analysis, we move more than 5% of teachers, so this comparison to teacher replacement policies might seem unbalanced. In Appendix Figure A7, we show that even replacing all below median teachers would achieve per-student gains of less than 0.054σ. Going the other way, in Appendix Figure A8 we show that reassigning just 10% of teachers delivers gains of over 0.02σ per student and full gains are nearly realized once 80% of teachers are reassigned.

In terms of distributional consequences, the first-best allocation entails larger gains for advantaged than for disadvantaged students. This finding is because (1) the current allocation slightly favors disadvantaged students and (2) the negative correlation between economic disadvantage and class sizes is more important than the positive correlation between comparative advantage in disadvantaged students and absolute advantage documented in Figure 2.

The overall gains come (1) from sorting teachers to schools based on comparative advantage and (2) from placing high absolute advantage teachers in schools with larger class sizes (see Appendix Figure A9). These assignments may be very different from the ones in the data, which introduces two concerns. First, if our "reassignments" are farther away than the in-sample variation we use to validate our value-added model, then we may be less confident in our output predictions. We find that while some teachers end up in classrooms with different composition or sizes from the ones where we observe them, this variation is still within the support of our data (see Appendix Figure A10). Second, by moving teachers across different student types, we are relying on the cardinality in the test score measures. As an alternative way to scale test scores, we express them in percentiles and find that the predicted gains from reallocation are nearly identical (see Appendix Table A5).

We include sorting based on class size for three reasons. First, it is a feature of the environment: there is class size dispersion and we show in Appendix C that this variation is persistent.[13] Second, teachers differ in their absolute advantage such that reassignments based on class size have the potential to matter for student achievement. Third, our validation exercise found that our output measures were mean forecast unbiased across class size changes. But because some readers

[13] In that Appendix, we also show that there are not systematic patterns of teachers "bargaining" over assignments within schools: i.e., we show that newly hired and more experienced teachers are not assigned smaller classes or fewer disadvantaged students within a school.

may prefer reassignments based solely on comparative advantage, we present the potential gains where we impose constant class sizes across all schools (see Appendix Table A6 for the full set of constant class size results). We estimate gains of 0.021σ, which are nearly identical to Delgado (2021)'s estimates using race-based match effects. These gains are 38% as large as the gains that incorporate class size variation. Thus, the potential to sort solely on comparative advantage remains economically meaningful.

How robust are our estimates of the efficiency gains to alternative value-added estimators? We find that the overall picture is quite robust across our three alternative estimators. The reason we find such large gains using the homogeneous value-added is that we are estimating fewer parameters: the dispersion in the "true" value-added is essentially identical in the homogeneous value-added as in the heterogeneous value-added, but there is less shrinkage and so the gains are larger.

Our specification likely misses some match effects.[14] We find similar results when we allow for match effects to vary with student race or prior academic achievement.[15] In Section 7.5 we show that our allocation conclusions are similar for these other forms of heterogeneity. Further, we conduct a simulation exercise where we allow our modeled form of match effects to be incomplete. We present the result in the top panel of Appendix Figure A11 and find that the potential gains only increase, such that our results may be a lower bound. We also find that our results from Section 7 are qualitatively unchanged.

Closing achievement gaps: If the district cared only about disadvantaged students, then it could achieve large gains for that type—specifically, it could close over a seventh of the achievement gap in levels between advantaged and disadvantaged students *in a single year*. The bottom panel of Table 3 shows that targeting non-disadvantaged students would yield a 0.075σ per-student gain, while leading to a 0.05σ per-student reduction for advantaged students.

How robust are our estimates of the equity gains to restricting to constant class size or alternative value-added estimators? The table shows that the gains to disadvantaged students and the losses to advantaged students are larger in all of our alternative specifications than in the baseline. Thus, our statements about the share of the achievement gap that could be closed are conservative.

4 The vacancy posting, application, and hiring process

We focus on the market for elementary-school classroom teachers for two reasons. First, teachers in these positions are typically responsible for instruction in the tested subjects and thus we can infer

[14] A form of potential match effects we have not included are same-race (between teacher and student) match effects (Dee, 2004, 2005; Gershenson et al., 2018), and same-gender match effects (Dee, 2005; Lim and Meer, 2017). In our data, we find minimal evidence of same-race or same-gender effects (see Appendix Table A7). A form of match effects we cannot test for in our data is that of teaching practices discussed in Aucejo et al. (2021) and Graham et al. (2020).

[15] See Appendix Table A8 for the structural parameters.

their quality from systematic gains in their students' test scores. Second, because these positions also have common certification requirements, we can reliably classify which teachers are eligible for the position.

4.1 Market overview

Our district organizes a decentralized hiring and transfer process in which teachers choose where to apply and principals choose whom to hire. External and internal (transfer) applicants are pooled into one market. Here we describe the basic market structure.

Market organization: The school district runs a centralized online hiring platform. Each school posts its openings on the platform, and teachers choose whether to apply to each posting.

Timing: We examine the "on-cycle" part of the market, which dictates hiring and transfers between school years. It begins in the winter, when the district notifies each school of known and expected attrition among the school's work force and of how many positions that school may hire. It ideally ends with filled positions by late August before the new school year. Similar to Papay and Kraft (2016), some schools are unable to fill all positions by the start of the new school year.

Postings: The number of postings at a school reflects a combination of enrollment, budget, and the number of teachers who leave. All three pieces of information are not necessarily known before the main hiring season starts. This information delay generates variation within and across schools in the timing of postings. For example, late information about enrollment or budget fluctuations often necessitate late posting. Or if there is mid-year attrition, then the school would know long before hiring season started that there would be a vacancy, which allows for early posting.

Applications: An application consists of a variety of documents including a teacher certification and a brief diversity statement. The same set of documents applies to all positions. Thus, a prospective teacher faces a fixed cost of preparing materials.

Evaluation and hiring: When a teacher applies to a position, the hiring school receives her application materials through the platform. The school's principal may then rate the applications and choose to interview applicants on a rolling basis. For known positions at the beginning of the hiring period, there is a short window during which only transfers from within the district are able to apply. Schools can either hire from this pool or wait and consider more applicants.

If the principal wants to hire the candidate, she extends a job offer. The candidate has 24 hours to accept the offer, and if the teacher accepts, she commits to not accepting an alternate offer in the same cycle.

4.2 Empirical features and implications for modeling teacher applications

We document eight features of the market that inform how we model it.

The first set of features leads us to treat teacher applications as non-strategic. The natural alternatives to non-strategic applications would be a portfolio choice problem (Chade, Lewis and Smith, 2014), possibly involving a waiting strategy. A portfolio choice problem would arise through positive marginal costs of each application or other interactions across applications. The following two features are inconsistent with these rationales:

Feature #1: The marginal cost of applications is essentially zero. Applying amounts to clicking a button that sends the standardized materials to the particular position. Indeed, a teacher certifies that she will not dis-intermediate the process. The website asks a teacher to sign the following statement: "I understand that I should not send materials to individual hiring managers or principals."

Feature #2: Principals do not see what other applications a teacher submits.

Some versions of waiting strategies amount to dynamic portfolio management, and so the previous two institutional features push against these being empirically relevant. More generally, a waiting strategy would be sub-optimal in the sense that a teacher could miss many potentially desirable vacancies because of the following feature:

Feature #3: Posting, applications, and hiring happen on a rolling basis throughout the hiring season. From April to August, both sides of the market operate on a rolling basis (Appendix Figure A12). The left columns in Table 4a show that the modal month for posting is June, with only 16% posting in April. The middle columns show that applications lag postings. The right columns show that hiring occurs on a rolling basis and tends to lag posting by about a month. Over half of hires are made by the end of June, even though over a quarter of positions have yet to be posted.

In practice, teacher application behavior appears inconsistent with a waiting strategy:

Feature #4: Teachers who are on the platform apply to vacancies very soon after they are posted. To characterize the timing of applications, we construct a measure of a teacher's wait time to apply to a vacancy. The wait time is the time elapsed between the first day a teacher could have applied to a vacancy and the day the teacher actually applied to the vacancy, where we assume that the teacher only learns that a vacancy is available on days she logs into the system and applies.[16]

[16]More formally, let $\mathcal{A}_{jt}$ denote the set of days where teacher j applied to at least one vacancy in year t, with $a_{jt} \in \mathcal{A}_{jt}$ measured in calendar days. Let b_{kt} be the (calendar) day that position k's vacancy is posted, and let c_{jkt} be the day that teacher j applies to position k. For every application j sent in year t, we define wait time w_{jkt} as: $w_{jkt} \equiv c_{jkt} - \min_{a_{jt} \in \mathcal{A}_{jt}: a_{jt} \geq b_{kt}} a_{jt}$.

Figure 3 shows that the wait time to apply for vacancies is typically very short. The top panel of Figure 3 shows that the median wait time to apply to vacancies that were already posted on the first day the teacher logged into the system (the "stock" of vacancies) is 0 days. Thus, the applicant's first day likely includes searching for older vacancies. Indeed, the mean vacancy an applicant applies to on day one has been posted for 23 days (Appendix Table A9). The bottom panel shows that the median wait time to apply to vacancies that were posted after the first day the teacher applies (the "flow" of vacancies) is also 0 days.

Feature #5: Teachers' application stopping behavior is hard to predict. In terms of applicants ending their search, many applicants' final applications come early enough in the cycle (9% in April or before, 16% in May, 22% in June) that they are potentially forgoing many yet-to-be-posted vacancies (Appendix Table A9, panel C). While some teachers who stop searching may have accepted a job, we see similar patterns among teachers who do not transfer that cycle. Thus, the end of search might be driven by shocks unrelated to accepting a job.[17]

How we model teachers' applications: These features lead us to treat applications as non-strategic and teachers' choice sets as all positions with postings active between a teacher's first and last application. In a full information environment, we would interpret the applications as revealing a teacher's preference for these vacancies. But if teachers were inattentive, then this inference would be mistaken. One empirical implication of inattention would be that teachers wait to apply to vacancies because they only notice the vacancy on the second or third (or nth) time that they use the platform. The absence of waiting is inconsistent with this implication of inattentiveness.

These assumptions imply very large applicant choice sets (Panel A of Figure 4), from which applicants apply to many positions (Panel B of Figure 4). These large choice sets and application sets allow us to precisely estimate heterogeneous preferences. For a case study of this heterogeneity, in Appendix D we present descriptive evidence of significant amounts of cross-teacher heterogeneity in application rates to Title I (high-poverty schools).

4.3 Empirical features and implications for modeling principal applications

We now turn to principals' choice sets, which we define as all of the applications they receive. Natural alternative assumptions include (1) due to rolling nature of the market, the position receives a meaningful number of applications after the principal has made a decision, or (2) because those teachers might still be in the market, the principal pays attention to the most recent applications.

[17] Teachers may continue searching after their final application day. The frequency of applications after the first application day is low enough that statistically we cannot rule out long periods of search without making an application. In Section 7.5, we report a robustness check of adding a seven day buffer to the end of the window.

The following feature is inconsistent with both of these alternatives (and is evidence against another strategic motive for teachers to time their applications):

Feature #6: The timing of applications that principals rate and do not rate is similar. We view all applications to each position. Table 5 shows that we see a hire in 80% of the postings. In 12% of postings, we see a declined offer. In 18% of postings, we see further principal evaluations and outcomes. We classify these into five groups: (1) interviews, (2) positive assessments, (3) neutral assessments, (4) negative assessments, and (5) application withdrawals.

From the data on the subset of vacancies for which we have multiple outcomes, the applications that receive ratings (i.e., a rating or an interview) have similar timing to those that principals do not rate. Table 5 shows the day of application relative to the application date of the eventual hire. Applications with principals' ratings are received on average only 2.2 days earlier than applications without ratings.

Our construction of choice sets implies that teachers and positions are not active for the whole cycle. With respect to identifying preferences, the concern would be that there is some systematic correlation between teacher and position characteristics and the timing of when they are active. Naturally, we cannot rule out all forms of sorting. Fact #5 speaks against teachers stopping their search strategically. We can also explore various forms of sorting based on observed characteristics. The following shows that there is little evidence for such sorting:

Feature #7: The timing of postings is hard to predict based on school characteristics. Institutionally, we have already discussed why posting—even within a school—is likely spread out: the arrival of relevant information is spread out.

One key source of heterogeneity in the timing of vacancies is that the the district has traditionally allocated replacement positions only once it is aware that a teacher is leaving, rather than "in expectation" of the number of vacancies. Since most teacher attrition occurs over the summer, this policy necessarily generates spread out posting. There are many reasons why teachers would not notify the school early enough for the school to post the job in April. For example, teachers may not know that they will leave until they have secured another position, setting up a vacancy chain in which schools that lose a transferring teacher must search later in the market. Or, teachers may withhold the information, particularly if they fear their leaving could negatively affect them.

While some of these factors suggest that there could be a systematic relationship between posting date and school type, we do not observe such patterns in the data. First, Table 4a shows that the months with highest shares of Title I postings occur early in the cycle (in April (62%) and May (52%)). This finding runs counter to a vacancy chain with Title I schools at the bottom. Second, there is vast variation in the timing of job postings within the same school. Table 4c pools posting dates across the years in our data and shows that 89% of schools that post jobs in July also post jobs in April. A similar pattern holds for schools with April postings.

Feature #8: The timing of applications is hard to predict based on teacher characteristics. We focus on one characteristic: the value-added of a teacher. Table 4b shows that teachers with above-median value-added scores apply slightly earlier in the cycle, but these differences are small.

How we model principals' actions: Principals take three actions: rating applicants, interviewing applicants, and hiring applicants. While it is conceivable that the offer decision might reflect strategic considerations (e.g., *is this teacher likely to accept the offer?*), such considerations are not relevant in the principal rating. We therefore use the principal rating as our primary indicator of principal preferences. (In our data, there is only a single field that records the principals' actions. If a positive assessment turns into an interview, then the field records an interview. Hence, we interpret the entry "interview" or "hired" as being an application that received a positive rating.)

In terms of the principal choice set, we view Feature #6 (the timing of applications that principals do and do not rate is similar) as suggestive that principals consider all applications. But we do not have additional evidence on the timing of principals' actions that would allow a more precise characterization of the process. In Section 7.5 we pursue a variety of robustness checks around this assumption.

5 Teacher preferences

5.1 Applications Model

We now formalize the discussion of how to infer teacher utilities from application choices given non-strategic applications. The district's labor market consists of a finite set of potential teachers, indexed by j, and a finite set of positions, indexed by p. Each position is associated with a specific school, $k = k(p)$, and may be assigned to at most one teacher. The exception is the outside option ($p = 0$), which includes leaving the district or teaching and has unlimited capacity.

At the beginning of year t, each teacher has an assignment, denoted by c. For teachers new to the district, this assignment is the outside option ($c = 0$), while for incumbent teachers, the assignment is j's position in the prior year, $c = p(j,t-1)$. Teachers may always keep their initial assignment. On an exogenous date $r = r(j,t)$, teacher j enters the transfer system.[18] If she enters, then she is active in the transfer system until an exogenous end date, $r' = r'(j,t)$.

If the teacher enters the transfer system, then she may apply to any position p that is active at some point between r and r'. There is no marginal cost to applying and there is no limit on the

[18] We treat the decision to enter the system as exogenous. We discuss selection into the system in Appendix E.

number of applications she can submit. Let a_{jpt} be an indicator for whether teacher j applied to position p in year t. Teachers' applications are known only to position p and teacher j.

These assumptions lead teachers to treat the application process non-strategically by applying to any position with utility higher than her current position and the outside option.[19] Slightly abusing notation (since $c = 0$ for teachers outside the district), a teacher submits an application to position p if:

$$a_{jpt} = \mathbf{1}\{u_{jpt} > \max\{u_{jct}, u_{j0t}\}\}, \tag{7}$$

where u_{jpt} is teacher j's utility from working at position p in time t.

5.2 Parameterization

We adopt a characteristics-based representation of teacher utilities over positions. By summarizing the position in terms of a lower-dimensional set of characteristics, we allow teachers to vary in their valuations of schools with these characteristics. Teacher utilities over positions are:

$$u_{jpt} = -\gamma d_{jpt} + \pi_j VA_{jpt} + \beta_j X_{pt} + \eta_{jt} + \varepsilon_{jpt}. \tag{8}$$

d_{jpt} is the one-way commute time (in minutes) between the teacher and the position and will serve as a numeraire for exposition (Appendix Figure A13 shows a binscatter of application probabilities against distance, revealing a strong downward slope until about 40 minutes). VA_{jpt} is teacher j's total value added at position p in year t.

Value-added, VA_{jpt}, combines absolute and comparative advantage. We define a teacher's absolute advantage to be her predicted value-added at a representative school: $AA_{jt} = n_{1t}\hat{\mu}_{j1t} + n_{2t}\hat{\mu}_{j2t}$, where n_{mt} is the average number of type m students in a classroom in the district. Comparative advantage, CA_{jpt}, at a specific position is then the difference between predicted value-added at school $k(p)$ and absolute advantage: $CA_{jpt} = VA_{jpt} - AA_{jt}$. Because we control for absolute advantage in the person-time effects, η_{jt}, the coefficient on VA_{jpt}, π_j, captures the strength of teachers' preferences for schools where their comparative advantage is high, reflecting the alignment between teachers' preferences and student output. We allow for preference heterogeneity by including a random coefficient in π_j:

$$\pi_j = \bar{\pi} + \sigma^{VA} \nu_j^{VA}, \tag{9}$$

where $\nu_j^{VA} \sim^{iid} N(0,1)$. Since π_j varies across teachers but we do not have random coefficients on absolute advantage, π_j includes both the preference over comparative advantage and any cross-

[19] We assume that any post-application steps necessary to be assigned to a position – e.g., interviews – are costless. In our data, teachers with multiple interviews are so rare that even if interviews are costly, they are rare enough that it is unlikely teachers consider dependence across applications.

teacher heterogeneity in preference over output.

X_{pt} is a vector of observed characteristics of positions: the fraction of a school's students that are economically disadvantaged (e), the fraction that are Black (b), the fraction that are Hispanic (h), and the fraction with an above median prior year math test score (s). We allow teachers to have heterogeneous preferences related to these school characteristics. Specifically,

$$\begin{aligned}
\beta_j^e &= \beta_{j0}^e + \beta_{j1}^e AA_{jt} + \sigma^e \nu_{jt}^e \\
\beta_j^b &= \beta_{j0}^b + \beta_{j1}^b AA_{jt} + \beta_{j2}^b Black_j + \sigma^b \nu_{jt}^b \\
\beta_j^h &= \beta_{j0}^h + \beta_{j1}^h AA_{jt} + \beta_{j2}^h Hispanic_j + \sigma^h \nu_{jt}^h \\
\beta_j^s &= \beta_{j0}^s + \beta_{j1}^s AA_{jt} + \sigma^s \nu_{jt}^s,
\end{aligned} \tag{10}$$

where $Black_j$ and $Hispanic_j$ are indicators for teacher race categories and ν_{jt} is a vector of independent, standard normal random coefficients. Thus, the σ parameters capture the standard deviation of idiosyncratic preferences related to each school characteristic.

We follow Mundlak (1978) and Chamberlain (1982) and model η_{jt} using correlated random effects. We model teacher-year unobserved heterogeneity in preferences for teaching in the district as the sum of several components:

$$\eta_{jt} = \lambda Z_{jt} + \rho CM_{jt} + \sigma^\eta \nu_{jt}^\eta. \tag{11}$$

Z_{jt} are teacher-year characteristics – whether the teacher is in the district, whether the teacher is Black, whether the teacher is Hispanic, whether the teacher is female, the teacher's predicted value-added for economically disadvantaged students, the teacher's predicted value-added for non-economically disadvantaged students, and dummy variables for whether the teacher has 2-3 years of prior experience, 4-6 years of prior experience, or more than 6 years of prior experience. CM_{jt} is a set of teacher-year averages of the variables that vary across the job postings within teacher-year (value-added, commute time, interactions of teacher and school characteristics). Thus, through CM_{jt}, we allow unobserved heterogeneity to be correlated with CA_{jpt} and X_{pt}. Finally, ν_{jt}^η is an independent standard normal random effect.

ε_{jpt} is an iid Type I extreme value error. Let $V_{jpt} = u_{jpt} - \varepsilon_{jpt}$ be j's representative value for position p in year t. Then the distributional assumption on ε_{jpt} implies that:

$$Pr(a_{jpt} = 1) = \frac{exp(V_{jpt})}{1 + exp(V_{jct}) + exp(V_{jpt})} \text{ and } Pr(a_{jpt} = 1) = \frac{exp(V_{jpt})}{1 + exp(V_{jpt})}, \tag{12}$$

for teachers already in the district and teachers new to the district, respectively.

5.3 Estimation and Identification

We estimate the teacher preference parameters using the teachers' applications to positions. We define a teacher's choice set, $\mathcal{P}_{jt}$, to be the set of vacancies active at the same time as the teacher. A teacher's start and end (search) date are the dates of her first and last application. Similarly, a vacancy's start and end (active) date are the dates it receives its first and last application.

We estimate teacher preferences via maximum simulated likelihood, where we simulate from the normal distributions of the random coefficients. Let n index each simulation iteration and let $A_{jptn}(\theta)$ be the model-predicted probability that j applies to position p in year t in simulation iteration n at parameter vector θ. For each teacher j in year t, we construct the simulated likelihood as:

$$L_{jt} = \frac{1}{100}\sum_{n=1}^{100}\prod_{p\in\mathcal{P}_{jt}}(a_{jpt}A_{jptn}(\theta)+(1-a_{jpt})(1-A_{jptn}(\theta))), \tag{13}$$

where a_{jpt} is an indicator for whether j applied to p in the data. Our full simulated log likelihood function is:

$$l = \frac{1}{J}\sum_{j} logL_{jt}. \tag{14}$$

In Section 4 we argued that it is empirically plausible that teachers apply non-strategically. Under this assumption, the choices that teachers make identify preferences and preference heterogeneity. Heuristically, if within her choice set a teacher is more likely to apply to positions with a particular characteristic than a position without this characteristic, then we infer that the teacher has a preference for schools with this characteristic. For mean coefficients, the relevant features of the data are the mean application rates to schools with certain characteristics. For observable preference heterogeneity, they are the variation in application rates across teacher characteristics. Finally, for unobservable preference heterogeneity, they are correlations at the teacher-level in the characteristics of the positions they apply to beyond what we would predict based on observables.

We seek to predict teachers' valuations over positions rather than causal effects of changes in position characteristics on choices. This still allows us to make counterfactual predictions. In Section 7, we make predictions by directly changing teachers' preferences. Mechanically, we assume that teachers only value output, rather than using estimated preferences. Conceptually, this combines a monetary bonus for output with information about output in each position, and thus only assumes that teachers place non-zero value on money. In Section 8, we give teachers bonuses as a function of school characteristics. Again, these counterfactuals do not require causal effects of characteristics in teacher preferences. For example, even if teachers place zero value on output, it is still possible to give bonuses for output by giving extra utility in proportion to the extra amount of output. We cannot consider counterfactuals where we change a particular characteristic. For example, we cannot use our estimates to predict the effects of changing the share of disadvantaged students in a school.

As a convenient way to interpret magnitudes, we sometimes convert utility to minutes of commute time, which requires the stronger assumption that commute time is exogenous. But because we primarily make relative comparisons of the costs of various policies, we do not rely on having consistently estimated the causal effect of commute time, unless noted.

5.4 Teacher Preference Estimates

Table 6 presents the teacher preference model estimates. First, consistent with what we found in simple summary statistics above, teachers prefer positions with greater shares of advantaged students. Second, teachers dislike positions with longer commutes. Finally, teachers have only slight preference toward positions where they have higher value-added.

Responsiveness to school and match characteristics varies with observable and unobservable heterogeneity. For example, teachers with higher absolute advantage have more negative preferences over the school's fraction of students that are disadvantaged. We also find a large positive same-race premium for Black teachers and schools with large fractions of Black students. In terms of unobservables, we typically find substantial dispersion in the random coefficients. For example, a standard deviation of the random coefficients on comparative advantage and fraction disadvantaged are each about 1.5 times the mean valuation.

To help interpret the strength of—and heterogeneity in—some of these relationships, Panels (a) through (c) of Figure 5 show how the average rank of positions in teachers' preferences change as single characteristics change, as well as the 10th and 90th percentile of these positions in teachers' rankings. We do not hold other characteristics fixed so that, for example, when we study commute time, other characteristics of schools are potentially changing. The figure emphasizes that commute time is a powerful predictor of rankings: changing commute time from 5 minutes to 25 minutes decreases the average rank of a position (for the average teacher) from about the 80th percentile to the 50th percentile. Similarly, the fraction of students that are disadvantaged is a powerful predictor of ranking: across the support, the mean ranking moves by about 20 percentiles, and if teachers were given their top choice there would be oversupply toward economically advantaged students (Appendix Figure A14). In contrast, while teachers do pursue comparative advantage, this relationship is quite weak: across the support of the data, varying teachers' comparative advantage only increases the rank of a position by a couple percentiles. The figures also emphasize that there is substantial heterogeneity in teachers' rankings of positions: across the support of most of these characteristics, the range from the 10th percentile in the teacher distribution to the 90th is very large.

Teachers' preferences are not particularly aligned with the first-best allocation that maximizes student achievement or that maximizes the achievement of disadvantaged students. Panel (a) of

Figure 6 shows that the mean ranking of the first-best position in teachers' preferences is the 48th percentile (we use the same sample as in Section 7). Even if on average teacher preferences do not align with the planner, stronger teachers having more aligned preferences could limit the misallocation resulting from teachers' preferences. The figure shows that this possibility does not occur: for the average strong teacher, the first-best position remains below her 50th percentile ranking. Panel (b) of Figure 6 makes a similar point for the objective of reducing achievement gaps: the rank of the allocation that minimizes the achievement gap hovers around the 50th percentile across the absolute advantage distribution, and if anything the rank is slightly decreasing in absolute advantage. Thus, giving teachers more choice might not produce efficiency or equity gains.

6 Principal behavior

6.1 Model and parameterization

Each position p is associated with a principal with the same index. Principal p derives non-wage utility $\tilde{v}_{jpt}$ from teacher j holding the position in year t. Because principals in our empirical context do not have to pay teacher wages out of a school budget, we model a principal as giving teacher j a positive rating ($b_{jpt} = 1$) if the non-wage utility is positive: $\tilde{v}_{jpt} > 0$.

We adopt a characteristics-based model and parameterize $\tilde{v}_{jpt}$ to be a linear function of position and teacher characteristics, a random effect, and an idiosyncratic teacher-position error:

$$\tilde{v}_{jpt} = \alpha_p W_{jpt} + \sigma_\kappa \kappa_{pt} + \upsilon_{jpt}. \tag{15}$$

To allow principal behavior to possibly align with output, W_{jpt} includes j's total value-added at school $k(p)$. We further include common teacher characteristics: teacher prior experience (in bins of 2-3 years, 4-6 years, and 7+ years), whether the teacher has a Masters degree, whether the teacher is Black, whether the teacher is Hispanic, and whether the teacher is female.[20] Finally, we include a constant and interact whether the teacher is Black with the fraction of the school's students that are Black and whether the teacher is Hispanic with the fraction of the school's students that are Hispanic. We allow principals to have heterogeneous valuations over teachers based on W_{jpt} by letting α_p vary with whether the school has Title I status.

To capture principals' heterogeneous outside options and variation in propensity to assign ratings, κ_{pt} is a normally distributed random effect. Finally, υ_{jpt} is i.i.d. Type I extreme value.

[20] We also include indicators for whether each demographic covariate is missing.

6.2 Identification

As with teachers, the identification of principal behavior is straightforward given our characterization of the process in Section 4. We observe the set of applications that a principal receives and we observe whether or not a principal gives an application a positive rating. We interpret the decision to give an application a positive rating as a non-strategic and costless action. This interpretation allows us to infer principal behavior from their choices in a straightforward way: those that are rated positively are preferred to those that are not. Because we observe the ratings, even if *interviewing is costly* and so principals are strategic at this stage, then our identification approach is still valid. One might also worry that assigning a rating is costly, and so it is done strategically. To alleviate this concern, we show below that if we restrict attention to applications where a principal assigned a rating (either positive or negative), then our results are quantitatively identical.

6.3 Estimates

To help understand the determinants of the principal ratings, Appendix Table A10 presents the changes in pseudo-R^2s from including different sets of observable teacher characteristics. The central finding of the table is that the main set of characteristics that explain ratings decisions are the combination of experience, licensing, certification and Praxis scores, which we showed in Table A2 poorly predict value-added. Value-added by itself or in addition to other characteristics generates very small changes in model fit.[21] More generally, observables have limited explanatory power for the ratings decision—the most saturated model has a pseudo-R^2 of less than 0.03.

Despite the small explanatory power of value-added in principal decisions, Table 7 shows that principals favor teachers with higher value-added.[22] Title I school principals rate Black and Hispanic teachers more positively than non-Title I teachers. Because Title I schools have a larger share of Black students, schools with higher fractions of Black students assign higher total valuations to Black teachers.

To help interpret the strength of the value-added relationship, Panel (d) of Figure 5, shows that the mean percentile of teachers in principals' ratings goes from the 25th percentile to the 60th percentile across the support of projected value-added. How much of this relationship is about value-added or the correlates of value-added? Consistent with the idea that observed characteristics poorly predict value-added, Appendix Figure A15 shows that if we omit value-added from the principal model then the relationship dramatically flattens.

How aligned are principals' ratings with the allocation that maximizes student achievement or that maximizes disadvantaged students' achievement? Panel (c) of Figure 5 shows that the average percentile of the output-maximizing teacher for a principal is the 52nd (compared to the 48th for

[21] EVAAS, the value-added measure computed by the state of North Carolina, has even less explanatory power.

[22] See Appendix F for the likelihood, which closely parallels the one for teachers.

teachers). Panel (d) shows a fairly similar set of results. Thus, schools' preferences are not very aligned with planner's preferred allocation, whether the planner has equity or efficiency objectives.

What do these relationships imply for the current allocation? Conceptually, we have shown that value-added matters but so do many other factors. Further, the mean valuation of a teacher with a characteristic (like high value-added) may not predict hiring choices well if schools can regularly hire the teachers at the top of their rankings. Thus, the role of principal behavior in generating the current allocation is non-obvious. The benefit of the equilibrium model is that we can precisely quantify the consequences of the observed structure of principal behavior relative to counterfactual structures.

7 Main results

We combine our estimated match-specific output from Section 3, our estimated teacher preferences from Section 5, our estimated principal valuations from Section 6, and our estimated market timing from Section 4 to simulate the market equilibrium. We first explore why there is parity in teacher quality between student types in the current allocation. We then consider trade-offs a policy-maker faces between teacher utility and student achievement when maximizing efficiency and the effects of various policies.

7.1 Simulation details

We make several choices in how we simulate allocations.

Sample: As elsewhere in the paper, we look only at non-specialized elementary school teachers and the equivalent positions.

We are interested in how to allocate the existing teacher workforce that is available to a district. While in Section 3, we considered all teachers within the district, for this exercise we consider allocating the set of *teachers who apply for positions* in the district in a given cycle, including teachers who are not currently in the district.

Because we want to be able to compute the output associated with each match, we restrict attention to the teachers for whom we can compute value-added, which includes teachers who have previously taught anywhere in the state. This restriction drops a large number of teachers: in the labor market for elementary school teachers in the 2015-2016 school year, we end up with 178 teachers and 296 positions.

Ashlagi, Kanoria and Leshno (2017) emphasize that imbalance in the sizes of the two sides of the market can determine the surplus accruing to each side of the market. To avoid the possibility of imbalance playing a role in our estimates, in each simulation run we randomly drop positions so that there are the same number of teachers and positions. In practice, we have also explored many of our results without dropping positions and we found that our results are very similar.

Randomness: While we estimate a distribution of random coefficients, in simulations we use a single draw of the random coefficients per teacher and principal. This draw is the one that maximizes the likelihood for the teacher or principal. We draw the type I errors in the preferences in an i.i.d. fashion.

Preferences: In using DA to find stable allocations, we have teachers and principals submit rankings according to their true preferences. If there are multiple equilibria, then for one side of the market it is not a dominant strategy to report truthfully. Below we show, however, that the equilibrium is essentially always unique and so truthful reporting is a dominant strategy.

Market clearing: We do not include an outside option when we run DA. Hence, given that we impose balanced markets, in each simulation all teachers are hired and all positions are filled.

To average over the randomness in both the errors and the random dropping of vacancies, we average over 200 simulation runs.

7.2 Model fit

We begin by considering the model's fit under status quo policies. We model the status quo as the teacher-propose equilibrium with restricted timing, and use estimated teacher and school preferences. Figure 7 shows that the model matches the basic qualitative patterns in the data: schools with a larger share of disadvantaged students have teachers (a) with stronger absolute advantage, (b) with comparative advantage in teaching economically disadvantaged students, (c) less likely to be experienced, and (d) more likely to be Black. Quantitatively, the model almost exactly matches the slope for teacher experience and whether teachers are Black. The model slightly underpredicts the slope in absolute advantage and misses some of the intercepts. The intercept gap comes from slight differences in the data and model samples (see Appendix E).

To assess whether our model fits better than alternate equilibrium assumptions, we examine the fit of models where schools and teachers match according to serial dictatorships. We find that a teacher serial dictatorship ordered by absolute advantage (Appendix Figure A16) and experience (Appendix Figure A17) and a principal serial dictatorship ordered by fraction of students that are economically disadvantaged (Appendix Figure A18) each produce a much worse fit than our model.

7.3 Understanding the current allocation

The top panel of Figure 8 shows the mean achievement of economically advantaged and disadvantaged students in various allocations. In the status quo allocation, we find that disadvantaged students have better teachers (the gap is slightly larger than 0.01 standard deviations).[23] Thus, the model reproduces the basic puzzle we have emphasized: the allocation is favorable towards disadvantaged students despite the fact that average teacher preferences are against working at such schools.

In terms of explanations for why the allocation is favorable to disadvantaged students, we find little role for market institution explanations. Changing the equilibrium from the teacher-proposing equilibrium to the school-proposing equilibrium has no effect on the allocation. Changing timing so that all vacancies and teachers are active at the same time increases output by a similar amount for both types. Thus, the current timing is not especially favorable to disadvantaged students.

We do find a large role for principal behavior. When we give principals output maximizing preferences, we end up in the allocation we might have expected based on average teacher preferences: now, advantaged students receive (quite dramatically) higher value-added teachers than disadvantaged students. Changing the principal preferences generates a change in the disadvantaged-advantaged student gap of 0.09σ in the quality of their teachers.

We find little role for teacher preference heterogeneity. We show this finding in two ways. First, we replace the estimated teacher preferences with homogeneous preferences by dropping all of the random coefficients from the preference specification. Going from the status quo to these homogeneous preferences barely changes the allocation. Second, if we have principals maximize value-added and teachers maximize value-added, then we return to an allocation that is favorable to disadvantaged students, which shows that it is teacher preferences that are the cause of the allocation that is favorable to advantaged students when principals maximize value added. (Table 8 shows that if teachers just maximized value-added and principals behaved according to their estimated preferences then the allocation is similarly favorable to disadvantaged students).

Why does principal behavior generate a distributionally favorable allocation; or, put differently, why does making principals rank according to output generate such unfavorable allocations? Figure 9 shows that in both the status quo and when principals only value output strong teachers end up in their more preferred schools. But this relationship tightens when principals only value output because it generates an allocation that is close to a serial dictatorship based on teacher absolute advantage. As we have shown, (strong) teachers tend to prefer teaching at schools with advantaged students. So the strongest teachers end up assigned to schools with advantaged students.

[23] The distribution in the status quo is slightly different than in the full sample. Like in the full sample, in the transfer sample class size is negatively correlated with the fraction of economically disadvantaged students and teachers with absolute advantage tend to have comparative advantage with economically disadvantaged students (Appendix Figure A19). But in the transfer sample, these factors balance out such that the first-best allocation splits the strongest teachers (Appendix Figure A20) and produces equal value-added across student types.

7.4 Trade-offs and counterfactuals

7.4.1 Trade-offs

The bottom panel of Figure 8 presents our main efficiency results. The production possibilities frontier (PPF) comes from solving a set of first-best problems (equation (3)) where we place different relative weight on students' achievement and teachers' utility. The top-left point shows the allocation of teachers to schools that maximizes student achievement. The bottom-right point show the allocation that maximizes teacher utility. There are two notable features of these points. First, there is a large gap in student achievement between the teacher and school first-best: the difference is 0.03 standard deviations of test scores. Second, there is a large gap in teacher utility between these allocations: the difference is about 35 minutes of one-way commuting time a day. A sensible valuation of an hour of commute time is about half of the hourly wage (Johnston, 2021). Hence, this finding, plus a causal interpretation of the commute time coefficient, implies that the gap between the teacher and school first-best is worth about one-sixteenth of a teachers' annual earnings.

The PPF also shows very favorable trade-offs available between teacher utility and student achievement. Concretely, if we start from the teacher-preferred allocation, then there are large gains in student achievement that barely affect teacher utility.

7.4.2 Counterfactuals

Turning to stable allocations, we study counterfactual policies, which parallels some of the exercises we previously saw. First, we find no role for equilibrium selection: changing from teacher-proposing to school-proposing DA has essentially no effect on the allocation. Second, policies that complete choice sets achieve 15% of the total allocative gains.

Third, making principals only value output—which combines an information intervention (telling principals teacher value-added) and an incentive—slightly **reduces** student achievement. This finding might appear counterintuitive as we are aligning principals' preferences with those of the planner and in Panel (e) of Figure 5 we showed considerable misalignment. Instead, the result reflects natural "theory of the second-best" reasoning and thus highlights important interactions between teacher and principal preferences. As we emphasized above, in the current allocation both principal and teacher preferences depart from the what the planner would want, and so fixing one in isolation can make the allocation worse.

Fourth, making teachers also only value output in a way analagous to principals in the previous paragraph has large effects on student achievement. If teachers only value output, then we achieve 74% of the total allocative gains available in this sample. We show below in Table 9 that there are similar gains if we change teacher preferences but do not change principal preferences.

Finally, once we complete choice sets and make both teachers and principals only value output, the remaining 26% of allocative gains is due exclusively to the absence of flexible prices. Prices play two roles in improving allocations. First, prices let the district change the agents' value from a match to align with that of the planner. Second, prices make utility transferable and allow the district to incorporate comparative advantage into principal preferences. Principals who only value output will rank teachers largely based on absolute advantage rather than comparative advantage.

Teacher utility in various allocations: Teacher utility increases as we move from the status quo and first expand choice sets and then make principals only value output. Each of these steps increases teacher utility on average by about 5 minutes of one way commute time. In contrast, if we make teachers only value output, but still evaluate the utility of the assignment using our estimated preferences, then we find that this change reduces teacher utility by about 20 minutes of one way commute time relative to the status quo.

7.5 Robustness

In Tables 9 and 8, we consider robustness along many dimensions.

1. Year of analysis: use the other years in our data;
2. Split of students: instead by race (white and non-white) and lagged achievement;
3. Teacher choice sets: add a seven-day buffer at the end, focus only on vacancies that were available on the first day the teacher applied;
4. Drop teachers who applied to only one vacancy;
5. Alternative preference models: various combinations of teacher and school fixed and random effects, as well as a correlated random coefficient specification for teachers;
6. Principal choice sets: restrict to applications submitted within plus or minus two weeks of the "hired" application; split the position-specific window when teachers were submitting in half and estimate separately on each half;
7. Use information about principal behavior differently: estimate a rank order logit model where we let, e.g., "hire" to be a better outcome than "positive assessment," use only applications where principals made an active choice (drop unrated applications), use a binary logit with hired as the outcome;
8. Hold class size fixed, rather than exploiting class size differences;
9. Value-added model: use the three alternative models previously discussed (homogeneous value-added, residualize differently, and use school means to compute fixed effects).
10. Principal model: add additional variables to the principal model.

Looking across Table 9, the magnitude of the efficiency gains of moving to the first best is quite robust across alternatives (if anything, our baseline tends to be conservative), and the basic pattern that making principals only value output reduces output is also robust. The one configuration where results are quantitatively, though not qualitatively, different is for constant class size where the gains are substantially smaller, as we might anticipate from Table 3.

Table 8 shows that the main distributional result is extremely robust across specifications: having principals maximize value-added has negative distributional consequences. Thus, the key explanation for the current equal allocation is principals not trying to hire the strongest teachers.

8 Teacher bonus counterfactuals

In the last section, we performed a structural decomposition and the quantified effects of idealized policies and found that the key change that would generate large increases in output is to align teacher preferences over schools with the output they would produce. We also found that total teacher utility drops considerably for the allocation that increases achievement, which suggests that teachers might require large compensation to accept the assignments. We now consider the effect of more realistic teacher bonus policies, similar to those that some districts have piloted. We emphasize important interactions across policies: the effects of teacher bonus policies depend on policies affecting principals.

8.1 Implementation details

We implement the bonuses as position-specific compensation, which importantly *does not* rely on having estimated causal effects of school characteristics in Section 5. The district offers a two-part bonus on the basis of a teacher-position characteristic, z_{jpt}, where each teacher receives b_0, a lump-sum amount, and $b_1 z_{jpt}$, a bonus b_1 per unit of z_{jpt}. Teacher j's utility for teaching at position p in year t thus becomes:

$$u_{jpt} = \tilde{u}_{jpt} + \gamma(b_0 + b_1 z_{jpt}), \tag{16}$$

where we multiply by the commute time coefficient (γ) to express bonus spending in minutes of commute time. We consider a range of b_1 for each z_{jpt}, which allows us to trace out the effects of different bonus sizes. For each b_1, we solve for the teacher-optimal stable equilibrium assignments, where $p^*(j)$ is j's assigned position, given the bonus size and the object that generates the bonus. Thus, because we give teachers utility directly for the characteristic, we do not use our estimated coefficients on the characteristics. The only coefficient we use is the one on commute time, which allows us to place our estimates in the same units as we used in Section 7.[24]

To focus on policies that are likely to receive teachers' support, we hold teachers harmless by making each teacher weakly better off than in the status quo equilibrium. Let $\Delta u_{jpt}^{b_1} = (\tilde{u}_{jp^*(j)t} - \tilde{u}_{jpt}) + \gamma b_1 z_{jp^*(j)t}$ be the change in teacher j's utility (excluding the transfer) between the zero-bonus and the b_1 bonus equilibria. We set the transfer such that the teacher with the worst change is indifferent:

$$b_0 = -\min_j \Delta u_{jpt}^{b_1}. \tag{17}$$

This lump-sum transfer can be either positive or negative, and so the district can pay teachers to enter this policy. Thus, the district's total cost to the bonus scheme is $b_0 + b_1 z_{jp^*(j)t}$, which depends on both the choice of b_1 and how it changes the allocation.

We examine bonus schemes over three objects (z_{jpt}). We start with bonuses for predicted output ($\sum_m n_{k(p)m} \hat{\mu}_{jm}$). Then we look at bonuses based on the fraction of disadvantaged students the teacher has ($p_{k(p)1t}$). These bonuses mimic the hard-to-staff school bonuses that some districts have piloted. Finally, we interact school and teacher characteristics by considering bonuses based on a teacher's absolute advantage times the fraction of disadvantaged students ($(p_{0t}\hat{\mu}_{j0t} + (1-p_{0t})\hat{\mu}_{j1t})p_{k(p)1t}$).

8.2 Results

Panel (a) of Figure 10 shows the effect of these three bonus schemes on overall achievement relative to the status quo. For reference, the top line shows the level of achievement in the the first-best allocation, the middle line shows the "best case" for bonuses when teachers and schools only value output, and the lower line shows the gains from simply changing market timing. To allow for comparisons across bonus schemes, the horizontal axis is the total realized spending (normalized to be in minutes of commute time per teacher).

The first notable aspect of this figure is that bonuses are more costly than the first-best policies depicted in Figure 8. Even at 150 minutes of (one way) commute time per teacher, bonuses still do not achieve the maximal student achievement, whereas we move from student to teacher-optimal policies with about 20 minutes of commute time. This large difference in cost is driven by the uniformity of the bonus scheme. Prices that implement the first best allocations take into account preference variation in a way that keeps costs down. For example, if a school is trying to convince a close-to-indifferent teacher to take a position, then the school only needs to increase the wage offer slightly for the teacher to accept the offer. We demonstrate the savings from flexible prices by allowing for separate lump sum payments to each teacher and plotting the gains in the dashed black line. At the spending level where the full potential gains are realized, the uniform bonus

[24]Unless noted, we will compare the effectiveness of bonuses with equivalent utility costs. Because we use the same conversion factor for all schemes, it does not affect the comparisons.

schemes have barely increased achievement. Second, paying directly for predicted achievement is the most efficient bonus scheme. In the status quo, disadvantaged students already have slightly better (matched) teachers and so paying teachers to be at schools with more disadvantaged students hardly increases output.

In Panel (b) we show the effects of the same bonus schemes on the difference in achievement between disadvantaged and advantaged students. We find that the predicted achievement bonuses and the bonuses for teaching disadvantaged students both lead to a widening of the baseline achievement gap, with the latter doing worse. This finding may be counter-intuitive, but the increased applicant pool for schools with disadvantaged students does not help these schools if they hire noisily. The bonuses targeted toward the best teachers teaching in disadvantaged schools lead to decreases in the achievement gap.

Consistent with the broad theme of this paper, the bottom panels show important interactions between principal preferences and the effectiveness of teacher bonuses. We conduct an identical exercise except that we pair the teacher bonuses with a bonus to principals such that principals only value output. As we have seen, when we pay principals for output, the distributional consequences change dramatically. Now, the fact that teachers have strong preferences against teaching at schools with larger shares of disadvantaged students means that there is a large range of spending where policies that target this issue directly are the most cost effective (panel (c)), while also lowering the achievement gap (panel (d)). As we saw in Section 7, achievement is higher when there is some force pushing back on teachers' preferences toward advantaged schools. In the absence of principals' heterogeneous valuations, bonuses targeted toward disadvantaged students serve this purpose.

Thus, we find that conditional on principal behavior, there is no trade-off between efficiency and equity (closing achievement gaps). But whether conditioning teacher compensation on predicted output or teaching disadvantaged students is the better policy depends on whether principals hire noisily or based on a teacher's predicted value-added.

9 Discussion

In this paper, we study the equity and efficiency consequences of the within-district allocation of teachers to schools. We find that the current allocation is equitable but not efficient. We investigate several explanations for the equity and determine that it is driven by principals not selecting their most effective applicants. Following the theory of the second best, this noisy principal hiring also leads to more efficient allocations by pushing back on the ability of high quality teachers to sort to advantaged schools. To capture most achievement gains, however, requires changing how teachers rank schools. This suggests the use of teacher bonus policies, though we find that their optimal form still depends on principal behavior.

In terms of caveats, we have followed the dominant strand of the literature and assumed that the only relevant measure of teacher quality is their value-added (in math). To the extent that there are other dimensions of teacher quality, then the allocation might look less distributionally favorable. In our counterfactual analysis, we have held fixed the assignment of students to schools (e.g., Abdulkadiroğlu, Agarwal and Pathak (2017)) and the distribution of class sizes (e.g., Angrist and Lavy (1999); Hoxby (2000); Leuven, Oosterbeek and Rønning (2008)). We have also held teacher and principal non-wage utility fixed in counterfactuals. But changes in malleable school characteristics, either under direct policy control (e.g., principal's support of teachers (Dizon-Ross, 2020; Johnston, 2021)) or that change in equilibrium (e.g., teacher peer effects (Jackson and Bruegmann, 2009)) may be a substitute or complement to the policies we consider. Exploring these richer dimensions is likely a useful area for future work.

More broadly, this paper has demonstrated the value of using rich data to study the functioning of particular labor markets. Here, our data allows us to estimate the behavior of the main agents in the market, rather than relying on strong assumptions to infer these from data on the observed equilibrium. In so doing, we have arrived at surprising conclusions about the determinants of the equilibrium and the design of policies. Presumably, other labor markets would also benefit from such analysis.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak. 2017. "The welfare effects of coordinated assignment: Evidence from the New York City high school match." *American Economic Review*, 107(12): 3635–89.

Agarwal, Nikhil, Charles Hodgson, and Paulo Somaini. 2020. "Choices and Outcomes in Assignment Mechanisms: The Allocation of Deceased Donor Kidneys." National Bureau of Economic Research Working Paper 28064.

Angrist, Joshua D, and Victor Lavy. 1999. "Using Maimonides' rule to estimate the effect of class size on scholastic achievement." *The Quarterly Journal of Economics*, 114(2): 533–575.

Angrist, Joshua, Peter Hull, Parag A. Pathak, and Christopher Walters. 2021. "Race and the Mismeasure of School Quality." NBER Working Paper 29608.

Antos, Joseph R., and Sherwin Rosen. 1975. "Discrimination in the market for public school teachers." *Journal of Econometrics*, 3: 123–150.

Ashlagi, Itai, Yash Kanoria, and Jacob D Leshno. 2017. "Unbalanced random matching markets: The stark effect of competition." *Journal of Political Economy*, 125(1): 69–98.

Atteberry, Allison, and Sarah E LaCour. 2020. "Putting Performance Pay to the Test: Effects of Denver ProComp on the Teacher Workforce and Student Outcomes." *Teachers College Record*, 122(12): 1–48.

Aucejo, Esteban M, Patrick Coate, Jane Cooley Fruehwirth, Sean Kelly, and Zachary Mozenter. 2021. "Teacher Effectiveness and Classroom Composition: Understanding Match Effects in the Classroom."

Ba, Bocar, Patrick Bayer, Nayoung Rim, Roman Rivera, and Modibo Sidibé. 2021. "Police Officer Assignment and Neighborhood Crime." National Bureau of Economic Research Working Ppaer 29243.

Ballou, Dale. 1996. "Do public schools hire the best applicants?" *The Quarterly Journal of Economics*, 111(1): 97–133.

Banerjee, Abhijit, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune. 2013. "Marry for what? Caste and mate selection in modern India." *American Economic Journal: Microeconomics*, 5(2): 33–72.

Barbieri, Gianna, Claudio Rossetti, and Paolo Sestito. 2011. "The determinants of teacher mobility: Evidence using Italian teachers' transfer applications." *Economics of Education Review*, 30(6): 1430–1444.

Bates, Michael. 2020. "Public and private employer learning: Evidence from the adoption of teacher value added." *Journal of Labor Economics*, 38(2): 375–420.

Bau, Natalie. Forthcoming. "Estimating An Equilibrium Model of Horizontal Competition in Education." *Journal of Political Economy*.

Biasi, Barbara. 2021. "The Labor Market for Teachers under Different Pay Schemes." *American Economic Journal: Economic Policy*, 13(3): 63–102.

Biasi, Barbara, Chao Fu, and John Stromme. 2021. "Equilibrium in Teachers' Labor Markets with Flexible Pay." National Bureau of Economic Research Working Paper 28530.

Bobba, Matteo, Tim Ederer, Gianmarco Leon-Ciliotta, Christopher Neilson, and Marco Nieddu. 2021. "Teacher compensation and structural inequality: Evidence from centralized teacher school choice in Perú." *Working Paper*.

Bonhomme, Stephane, Gregory Jolivet, and Edwin Leuven. 2016. "School characteristics and teacher turnover: assessing the role of preferences and opportunities." *Economic Journal*, 126: 1342–1371.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and Jim Wyckoff. 2013. "Analyzing the determinants of the matching of public school teachers to jobs: disentangling the preferences of teachers and employers." *Journal of Labor Economics*, 31(1): 83–117.

Boyd, Don, Hamp Lankford, Susanna Loeb, Matthew Ronfeldt, and Jim Wyckoff. 2011. "The role of teacher quality in retention and hiring: using applications to transfer to uncover preferences for teachers and schools." *Journal of Policy Analysis and Management*, 30(1): 88–110.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics*, 36(S1): S13–S70.

Chade, Hector, Gregory Lewis, and Lones Smith. 2014. "Student portfolios and the college admissions problem." *Review of Economic Studies*, 81(3): 971–1002.

Chamberlain, Gary. 1982. "Multivariate regression models for panel data." *Journal of Econometrics*, 18(1): 5–46.

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

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*b*. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review*, 104(9): 2633–79.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. 2008. "Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina." *Journal of Public Economics*, 92(5-6): 1352–1370.

Combe, Julien, Olivier Tercieux, and Camille Terrier. Forthcoming. "The design of teacher assignment: theory and evidence." *Review of Economic Studies*.

Combe, Julien, Umut Dur, Olivier Tercieux, Camille Terrier, and M Utku Unver. 2021. "Market Design for Distributional Objectives in (Re) assignment: An Application to Improve the Distribution of Teachers in Schools." Working Paper.

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

Cowgill, Bo, Jonathan Davis, B Pablo Montagnes, and Patryk Perkowski. 2021. "Matching for Strategic Organizations: Theory and Empirics from Internal Labor Markets." *Available at SSRN*.

Dahlstrand, Amanda. 2021. "Defying Distance? The Provision of Services in the Digital Age."

Davis, Jonathan M.V. 2021. "Labor Market Design Can Improve Match Outcomes: Evidence from Matching Teach For America Teachers to Schools."

Dee, Thomas S. 2004. "Teachers, race, and student achievement in a randomized experiment." *Review of Economics and Statistics*, 86(1): 195–210.

Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" *American Economic Review*, 95(2): 158–165.

Delgado, William. 2021. "Heterogeneous Teacher Effects, Comparative Advantage, and Match Quality."

Dizon-Ross, Rebecca. 2020. "How Does School Accountability Affect Teachers? Evidence from New York City." *Journal of Human Resources*, 55(1): 76–118.

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

Fox, Lindsay. 2016. "Inferring school desirability based on teacher transfer requests in New York City." *Working Paper*.

Fox, Zach. 2017. "Cherokee schools offer $10K signing bonus to retain new teachers." *Go Upstate*. `https://www.goupstate.com/story/news/education/2017/01/19/cherokee-schools-offer-10k-signing-bonus-to-retain-new-teachers/22639001007/`. Last accessed: December 26, 2021.

Gershenson, Seth, Cassandra MD Hart, Joshua Hyman, Constance Lindsay, and Nicholas W. Papageorge. 2018. "The long-run impacts of same-race teachers." *National Bureau of Economic Research*.

Goldhaber, Dan, Lesley Lavery, and Roddy Theobald. 2015. "Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students." *Educational researcher*, 44(5): 293–307.

Goldhaber, Dan, Roddy Theobald, and Danielle Fumia. 2022. "The role of teachers and schools in explaining STEM outcome gaps." *Social Science Research*, 102709.

Graham, Bryan S., Geert Ridder, Petra M. Thiemann, and Gema Zamarro. 2020. "Teacher-to-classroom assignment and student achievement." NBER Working Paper 27543.

Hanushek, Eric A. 2009. "Teacher Deselection." *Creating a New Teaching Profession*.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "Why public schools lose teachers." *Journal of Human Resources*, 39(2): 326–354.

Harris, Douglas N, and Tim R Sass. 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics*, 95(7-8): 798–812.

Hinrichs, Peter. 2021. "What kind of teachers are schools looking for? Evidence from a randomized field experiment." *Journal of Economic Behavior & Organization*, 186: 395–411.

Hitsch, Gunter J, Ali Hortaçsu, and Dan Ariely. 2010. "Matching and sorting in online dating." *American Economic Review*, 100(1): 130–63.

Hoxby, Caroline M. 2000. "The effects of class size on student achievement: New evidence from population variation." *The Quarterly Journal of Economics*, 115(4): 1239–1285.

Isenberg, Eric, Jeffrey Max, Philip Gleason, and Jonah Deutsch. 2022. "Do Low-Income Students Have Equal Access to Effective Teachers?" *Educational Evaluation and Policy Analysis*, 44(2): 234–256.

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

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

Jackson, C Kirabo, and Elias Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." *American Economic Journal: Applied Economics*, 1(4): 85–108.

Jacob, Brian A, Jonah E Rockoff, Eric S Taylor, Benjamin Lindy, and Rachel Rosen. 2018. "Teacher applicant hiring and teacher performance: Evidence from DC public schools." *Journal of Public Economics*, 166: 81–97.

Jatusripitak, Napat. 2018. "Three Essays in Applied Microeconomics."

Johnston, Andrew C. 2021. "Teacher preferences, working conditions, and compensation structure." Working Paper.

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

Kraft, Matthew A, John P Papay, Leigh Wedenoja, and Nathan Jones. 2020. "The Benefits of Early and Unconstrained Hiring: Evidence from Teacher Labor Markets."

Laverde, Mariana, Elton Mykerezi, Aaron Soujourner, and Aradhya Sood. 2021. "School district wage structure and teacher mobility: Assessing teacher preferences from inter-school mobility in Minnesota." *Working Paper*.

Leuven, Edwin, Hessel Oosterbeek, and Marte Rønning. 2008. "Quasi-experimental estimates of the effect of class size on achievement in Norway." *Scandinavian Journal of Economics*, 110(4): 663–693.

Levin, Jessica, and Meredith Quinn. 2003. "Missed Opportunities: How We Keep High-Quality Teachers out of Urban Classrooms."

Lim, Jaegeum, and Jonathan Meer. 2017. "The impact of teacher–student gender matches random assignment evidence from South Korea." *Journal of Human Resources*, 52(4): 979–997.

Lipsey, Richard G., and Kelvin Lancaster. 1956. "The general theory of second best." *The review of Economic Studies*, 24(1): 11–32.

Mansfield, Richard K. 2015. "Teacher quality and student inequality." *Journal of Labor Economics*, 33(3): 751–788.

Marcotte, David E. 2015. "Performance pay and the impact on teacher motivation." PhD diss. Indiana State University.

Mundlak, Yair. 1978. "On the pooling of time series and cross section data." *Econometrica*, 69–85.

Neal, Derek. 2011. "The Design of Performance Pay in Education." In *Handbook of the Economics of Education*. Vol. 4, 495–550. Elsevier.

Oyer, Paul, and Scott J Schaefer. 2011. "Personnel economics: Hiring and incentives." *Handbook of Labor Economics*, 4(PART B): 1769–1823.

Papay, John P, and Matthew A Kraft. 2016. "The productivity costs of inefficient hiring practices: Evidence from late teacher hiring." *Journal of Policy Analysis and Management*, 35(4): 791–817.

Pridemore, Adam. 2017. "State budget creates new principal pay plan and bonus programs." *North Carolina School Superindents Association*. `https://www.ncasa.net/site/default.aspx?PageType=3&DomainID=8&ModuleInstanceID=9&ViewID=6446EE88-D30C-497E-9316-3F8874B3E108&RenderLoc=0&FlexDataID=2754&PageID=9`. Last accessed: December 26, 2021.

Rivkin, Steven G, Eric A Hanushek, and John F Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica*, 73(2): 417–458.

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

Rockoff, Jonah E, Douglas O Staiger, Thomas J Kane, and Eric S Taylor. 2012. "Information and employee evaluation: Evidence from a randomized intervention in public schools." *American Economic Review*, 102(7): 3184–3213.

Rosen, Sherwin. 1986. "The theory of equalizing differences." *Handbook of labor economics*, 1: 641–692.

Roth, Alvin E., and Marilda A. Oliveira Sotomayor. 1992. *Two Sided-Matching: A Study in Game-Theoretic Modeling and Analysis.* Cambridge University Press.

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

Sass, Tim R., Jane Hannaway, Zeyu Xu, David N. Figlio, and Li Feng. 2012. "Value added of teachers in high-poverty schools and lower poverty schools." *Journal of Urban Economics*, 72: 104–122.

Shapley, Lloyd S, and Martin Shubik. 1971. "The assignment game I: The core." *International Journal of Game Theory*, 1(1): 111–130.

Sorkin, Isaac. 2018. "Ranking firms using revealed preference." *Quarterly Journal of Economics*, 133(3): 1331–1393.

Staiger, Douglas O, and Jonah E Rockoff. 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives*, 24(3): 97–118.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199.

Tincani, Michela M. 2021. "Teacher Labor Markets, School Vouchers, and Student Cognitive Achievement: Evidence from Chile." *Quantitative Economics*, 12(1): 173–216.

Table 1: Forecast Unbiasedness Tests for Value-Added Predictions

	Mean Res	Mean Diff	Mean Res	Mean Res	Mean Res	Mean Res
VA (Heterog)	1.052			1.060		
	(0.00650)			(0.00681)		
VA Diff		0.879				
		(0.0243)				
Post Transfer			-0.00243	0.00576		
			(0.00367)	(0.00280)		
VA * Post Transfer				-0.0885		
				(0.0212)		
VA – below 10th (disadv)					0.990	
					(0.0223)	
VA – 10th-90th (disadv)					1.058	
					(0.00698)	
VA – above 90th (disadv)					1.066	
					(0.0228)	
VA – below 10th (size)						1.011
						(0.0224)
VA – 10th-90th (size)						1.066
						(0.00713)
VA – above 90th (size)						0.961
						(0.0188)
Constant	0.00810	0.0477	0.00779	0.00745	0.00810	0.00800
	(0.000835)	(0.00101)	(0.00174)	(0.000883)	(0.000835)	(0.000843)
Subject	Math	Math	Math	Math	Math	Math
Mean DV	0.00764	0.0527	0.00754	0.00764	0.00764	0.00764
Clusters	21514	21514	21834	21514	21514	21514
N	74552	74552	75459	74552	74552	74552

The table includes tests of whether a value-added estimate is forecast unbiased. In the first and third through sixth columns, the outcome ("Mean Res") is the mean student math test score, residualized by student demographics including lagged scores, school fixed effects, and teacher experience measures. The mean is taken over all students for a given teacher-year. In the second column, the outcome ("Mean Diff") is the difference in the mean residualized math scores between a teacher's economically disadvantaged and advantaged students. The "VA" measures allow for match effects ("Heterog"). The measures predict mean student residuals using data from all prior years a teacher taught. "VA Diff" is the difference in predicted value-added between a teacher's economically disadvantaged and advantaged students (i.e., the predicted comparative advantage). "Post Transfer" refers to years after a teacher switched schools. The interaction with "VA" multiplies the post-transfer indicator with the heterogeneous value-added measure. Column (4) splits the year t observations into bins as a function of the change in share of disadvantaged students relative to the data observed for the teacher before year t. The split is based on percentiles of the change. Column (5) splits the year t observations into bins as a function of the change in classroom size relative to the data observed for the teacher before year t. The split is based on percentiles of the change. For columns (4) and (5) the p-value comes from F-test that the three coefficients are equal. Standard errors are clustered at the teacher level.

Table 2: Summary statistics for 2015-16, by economic disadvantage

	Focal, Adv	Focal, Disadv	Other, Adv	Other, Disadv
Students				
Male (%)	50.58	51.07	51.17	51.36
White (%)	64.61	9.11	75.58	35.09
Black (%)	17.04	51.78	9.54	32.63
Hispanic (%)	6.77	32.58	6.00	23.90
Other minorities (%)	11.58	6.53	8.88	8.38
Student performance (level scores)				
Math	0.70	-0.16	0.43	-0.30
Student performance (gain scores)				
Math	0.07	0.07	-0.01	0.00
Teachers				
Experience (% of teachers)				
0 years	4.32	10.99	3.34	4.85
1-2 years	10.45	17.24	6.90	9.80
3-5 years	17.33	19.31	11.22	12.83
6-12 years	29.47	22.99	26.72	26.18
13-20 years	21.98	18.72	28.57	24.40
21-27 years	9.77	4.34	12.53	11.43
28 or more years	6.68	6.40	10.73	10.50
Graduate degree (%)	45.20	43.28	39.66	37.44
Regular license (%)	97.10	87.17	97.84	94.71
NBPTS certified (%)	16.08	6.81	14.27	9.95
Praxis score	0.37	0.03	0.29	0.13
Attrition rate (%)				
From school	13.96	25.17	15.50	19.35
From district	8.44	13.62	10.21	13.09
Mean math value-added				
Baseline, econ disadv	0.02	0.02	-0.01	-0.00
Baseline, econ adv	0.01	0.02	-0.02	-0.01
Homogeneous	0.02	0.01	-0.00	-0.01
Using school means	0.16	0.13	0.08	0.09
Using alternative FEs	0.05	0.05	0.02	0.03

The table shows mean student and teacher in our sample for the 2015-16 school year. We split the sample into whether the student is in our focal district ("Focal") or in the rest of North Carolina ("Other") and whether he or she is economically advantaged ("Adv") or disadvantaged ("Disadv"). Math scores are standardized to have mean 0 and standard deviation 1 at the state-grade-year level. The alternate VA estimators are a (a) homogeneous value-added model with constant effects across student types, (b) a model that uses school mean characteristics rather than school fixed effects, and (c) a model that uses teacher-year fixed effects, rather than teacher-class-student type fixed effects, in the first residualization step.

Table 3: Potential Gains from Reassignment

	Per-Student Gains (σ)	As a Fraction of (Best-Actual)	Non-Disadvantaged	Disadvantaged
Alternate Allocations				
Best	0.054		0.095	0.018
Random	-0.003	-0.05	0.019	-0.023
Worst	-0.057	-1.06	-0.053	-0.062
Alternate Policies				
Best w/i School	0.013	0.28	0.015	0.010
Random w/i School	-0.000	-0.00	-0.000	-0.000
Replace Bottom 5% of Teachers	0.012	0.22	0.015	0.009
Targeting Disadvantaged Students				
Max Disadvantaged VA	0.016	0.30	-0.049	0.075
Best, Robustness				
Constant Class Size	0.021	0.38	0.018	0.023
Homogeneous VA	0.044		0.094	0.009
Using School Means in VA	0.053		0.079	0.030
Using Alternative FEs in VA	0.050		0.087	0.016
Max Disadvantaged VA, Robustness				
Constant Class Size	0.005	0.22	-0.111	0.096
Homogeneous VA	0.011	0.25	-0.081	0.076
Using School Means in VA	0.019	0.35	-0.056	0.084
Using Alternative FEs in VA	0.014	0.28	-0.055	0.076

The table shows the potential gains from reassignments of teachers to different schools. The sample is all teachers with non-missing value-added forecasts in 2016 (based on prior data), along with their corresponding 2016 assignments. Gains come from better matching of teachers to students, as teachers' effectiveness may differ across student types, and placing better teachers in schools with larger class sizes. The first column shows the per-student gains from various allocations relative to the actual allocation. Gains are measured in student standard deviations (σ). The second column shows the gain as a fraction of the full difference between the best (output-maximizing) and actual allocations. The third and fourth columns show the per-student gains, relative to the actual allocation, for non-disadvantaged and disadvantaged students. The best and random within school allocations only change the teacher-classroom assignments within a school. "Replacing Bottom 5% of Teachers" refers to replacing the bottom 5% of teachers according to realized per-student output with teachers with median value-added for each student type. The allocations that target particular student types maximize per-student output for students of one type only. "Constant Class Size" imposes an equal number of students (but possibly different composition) across all classes, in both the best and actual allocations. The alternate VA estimators are a (a) homogeneous value-added model with constant effects across student types, (b) a model that uses school mean characteristics rather than school fixed effects, and (c) a model that uses teacher-year fixed effects, rather than teacher-class-student type fixed effects, in the first residualization step. We assign classrooms the mean student composition and class sizes in that school in 2016 in all allocations except the "Best w/i School," "Random w/i School," and "Constant Class Size" allocations.

Table 4: Timing of posting, applying, and hiring

(a) Monthly shares by position

	Posting			***Applying***			***Hiring***		
	Vacs	Share	Share TI	Apps	Share	Share TI	Apps	Share	Share TI
April	295	16.24	0.62	24799	7.13	0.50	393	13.23	0.69
May	392	21.57	0.52	70248	20.21	0.50	585	19.70	0.63
June	502	27.63	0.52	108776	31.29	0.51	827	27.85	0.60
July	451	24.82	0.42	94171	27.09	0.50	755	25.42	0.50
August	167	9.19	0.46	44673	12.85	0.51	358	12.05	0.57
Total	1807	100		342667	100		2918	2918	

(b) Monthly shares by teacher value-added

	Has VA			***Above median VA***			***Top decile VA***		
	Apps	Share	Share TI	Apps	Share	Share TI	Apps	Share	Share TI
April	3050	6.23	0.44	1552	7.16	0.42	373	9.15	0.41
May	9662	19.75	0.44	4218	19.46	0.44	918	22.53	0.45
June	16832	34.40	0.46	8035	37.08	0.45	1396	34.26	0.47
July	13673	27.95	0.47	5600	25.84	0.46	944	23.17	0.46
August	5522	11.29	0.48	2189	10.10	0.47	434	10.65	0.52
Total	48739	100		21594	100		4065	100	

(c) Early vs. late posting times by school

	Posts in July		
Posts in April	No	Yes	Total
No	8	15	23
Yes	10	88	98
Total	18	103	121

This table shows the timing of posting, applying, and hiring during a cycle. Panel (a) shows the distribution of vacancy postings, applications, and hires by month, where hires correspond to the timing of the applicant who was hired to the position. For each type of action, we show the share that corresponds to Title I positions. Some of the vacancies produce multiple hires. In Panel (b) we show the distribution of applications by month, where we split the sample of applicants into those with a value-added forecast (i.e., had taught in tested grades and subjects in North Carolina prior to applying), those with above median value-added, and those in the top decile. Panel (c) shows the cross-tabulation of whether a school posts a vacancy in April and whether that school posts a vacancy in July (in the same cycle).

Table 5: Application evaluations, outcomes, and timing

	Hired successfully	Hired but taught elsewhere	Hired but not in district	Declined offer	Interview	Positive	Middle	Negative	Withdrew	No comment
mean	0.00051	0.00003	0.00017	0.00006	0.00000	0.00064	0.00029	0.00037	0.00002	0.07367
count	2,291	122	750	292	7	2,887	1,300	1,655	74	333,780

(a) Outcomes at the application level

	Hired	Declined offer	Interview	Positive	Middle	Negative	Withdrew	No comment	Any Non-Hire Action
mean	0.799	0.117	0.001	0.101	0.023	0.075	0.037	0.985	0.179
count	1,457	213	2	184	42	136	67	1,797	327

(b) Outcomes at the position level

	Obs	Mean	10th	25th	50th	75th	90th	Std. dev.
All applications	343,161	-0.0	-15.6	-5.8	-0.8	4.6	16.4	14.74
No notes	333,780	0.1	-15.2	-5.6	-0.7	4.5	16.1	14.38
Evaluated with notes	9,381	-2.0	-32.1	-15.0	-4.1	7.9	31.7	24.26

(c) Timing relative to hired applicant

This table shows the frequency and timing of application outcomes. The data record a single outcome per application; as an example, "Interview" implies not hired as otherwise the "Interview" outcome would be replaced by "Hired." The data record "Hired," which we split into "Hired successfully" for teachers who taught in the position's school the following year, "Hired but taught elsewhere" for teachers hired who taught in district but not at that position's school, and "Hired but not in district" for teachers hired who did not appear in the district the following year. "Positive," "Middle," and "Negative" reflect the authors' coding of different text categories. "No comment" includes applications without an updated status. Panel (a) shows frequencies at the application level and panel (b) shows frequencies at the position level for at least one outcome across all applications to that position (i.e., "Hired" indicates at least one application led to a hire). "Any Non-Hire Action" is a positive, middle, or negative assessment or an application withdrawal. In panel (c) we calculate the difference in timing (in days) between when an application was made and when the application that led to a hire was made. A value of 1 would indicate an application made 1 day after the one that led to a hire. In the last two rows, we split the sample into those with no notes ("No comment") and those with an outcome.

Table 6: Teacher preference estimates

	Estimate	Standard Error
Constant	2.032	4.453
Commute Time	-0.073	0.001
Commute Time Missing	-1.660	0.223
Value Added	0.081	0.008
St Dev Value Added RC	0.128	0.007
School Characteristics and Interactions		
Fraction Disadvantaged	-1.188	0.136
Fraction Black	-0.452	0.132
Fraction Hispanic	0.441	0.144
Fraction Above Median Achievement	0.163	0.149
Abs Adv x Fraction Disadvantaged	-0.797	1.029
Abs Adv x Fraction Black	-1.635	1.025
Abs Adv x Fraction Hispanic	2.487	1.074
Abs Adv x Fraction Above Median Achievement	-1.997	1.185
Black x Fraction Black	1.072	0.130
Hispanic x Fraction Hispanic	0.491	0.771
St Dev Fraction Disadvantaged RC	1.591	0.034
St Dev Fraction Black RC	1.296	0.054
St Dev Fraction Hispanic RC	0.637	0.065
St Dev Fraction Above Median Achievement RC	1.397	0.045
Teacher Characteristics		
VA Non-Disadvantaged Students	0.746	0.307
VA Disadvantaged Students	0.937	0.331
In District	-0.509	0.061
Black	-0.095	1.043
Hispanic	6.017	3.762
Female	0.284	0.064
Experience 2-3	0.070	0.083
Experience 4-6	-0.268	0.082
Experience 7+	-0.141	0.074
St Dev Random Effect	1.687	0.030
Chamberlain-Mundlak Device		
Fraction Disadvantaged Mean	-1.903	3.182
Commute Time Mean	0.032	0.004
Commute Time Missing Mean	1.231	0.249
Value Added Mean	-0.489	0.295
Fraction Black Mean	-2.786	2.707
Fraction Hispanic Mean	0.041	2.457
Fraction Above Median Achievement Mean	-0.986	4.718
Abs Adv x Fraction Disadvantaged Mean	-37.628	19.086
Abs Adv x Fraction Black Mean	36.183	18.362
Abs Adv x Fraction Hispanic Mean	15.838	19.942
Abs Adv x Fraction Above Median Achievement Mean	-16.346	6.488
Black x Fraction Black Mean	-2.200	2.412
Hispanic x Fraction Hispanic Mean	-20.462	14.686
Number of Students Mean	0.009	0.023

The table shows teacher preference coefficients, estimated using maximum simulated likelihood. We model the probability that a teacher applies to a position where the alternate options are not teaching in the district or keeping the current position. Random coefficients ("RC") are independent and simulated from the standard normal distribution. We model unobserved teacher-year heterogeneity using a Mundlak (1978) and Chamberlain (1982) device, taking the mean of each covariate across an applicant's choices. Commute time is measured in minutes, value added is total predicted output. Experience below 2 years is the omitted category.

Table 7: Principal preference estimates

	Estimate	Standard Error
Constant	-4.363	0.127
St Dev Random Effect	1.531	0.022
Title I	0.521	0.156
Value-Added	0.092	0.026
Value-Added x Title I	0.038	0.034
Experience 2-3	0.351	0.128
Experience 2-3 x Title I	-0.005	0.163
Experience 4-6	0.271	0.117
Experience 4-6 x Title I	0.035	0.160
Experience 7+	0.097	0.089
Experience 7+ x Title I	-0.344	0.120
Experience Missing	-0.342	0.060
Experience Missing x Title I	0.371	0.086
Masters	0.188	0.098
Masters x Title I	0.124	0.125
Black	-1.035	0.227
Black x Title I	1.722	0.453
Black x Fraction Black	0.396	0.267
Black x Fraction Black x Title I	-0.253	0.511
Hispanic	-0.690	0.454
Hispanic x Title I	0.450	0.561
Hispanic x Fraction Hispanic	2.259	2.219
Hispanic x Fraction Hispanic x Title I	-1.833	2.345
Female	0.053	0.106
Female x Title I	0.031	0.129
Gender Missing	-0.327	0.230
Gender Missing x Title I	-0.197	0.277
Race Missing	-0.530	0.210
Race Missing x Title I	0.374	0.247
VA Missing	0.490	0.089
VA Missing x Title I	-0.230	0.124

The table shows principal preference coefficients, estimated using maximum simulated likelihood. We model the probability that a principal submits a positive outcome (hire, interview, positive rating) for an application. Random effects are simulated from the normal distribution. Experience below 2 years is the omitted category. Value-added is total predicted output.

Table 8: Robustness: disadvantaged minus advantaged achievement

	Status quo	All Options	Principal Max VA	Teach Max VA	Both Max VA	First Best
Baseline	0.0204	0.0163	-0.0786	0.0088	0.0231	-0.0006
1. Vary year: baseline is 2016						
2012	-0.0169	-0.0198	-0.0943	-0.0388	-0.0030	-0.0454
2013	0.0023	-0.0028	-0.0885	-0.0109	0.0206	-0.0100
2014	0.0157	0.0104	-0.0903	-0.0229	0.0112	-0.0303
2015	0.0224	0.0216	-0.0819	0.0029	0.0633	-0.0020
2017	0.0351	0.0356	-0.0641	0.0247	0.0556	0.0187
2. Vary student type split: baseline is economic disadvantage						
Achievement	-0.0166	-0.0272	-0.1384	-0.0310	0.0109	-0.0321
Race	-0.0082	-0.0169	-0.1190	-0.0250	-0.0044	-0.0333
3. Vary choice set construction for teachers						
7 day buffer	0.0163	0.0112	-0.0887	0.0088	0.0231	-0.0006
First day choice sets only	0.0157	0.0232	-0.0836	0.0088	0.0231	-0.0006
Drop single app. teachers	0.0197	0.0125	-0.0813	0.0113	0.0289	0.0043
4. Vary teacher preference specification to use binary logit						
No REs or FEs	0.0178	0.0122	-0.0609	0.0105	0.0233	-0.0011
School FEs	0.0141	0.0098	-0.0042	0.0101	0.0261	-0.0006
School REs	0.0171	0.0116	-0.0660	0.0095	0.0260	-0.0005
Teacher FEs	0.0159	0.0119	-0.0679	0.0088	0.0231	-0.0006
Teacher REs	0.0187	0.0119	-0.0681	0.0089	0.0234	-0.0013
Teacher REs, School FEs	0.0179	0.0093	-0.0611	0.0097	0.0242	-0.0005
Teacher FEs, School FEs	0.0181	0.0096	-0.0604	0.0088	0.0245	-0.0006
5. Allow for correlated random coefficients in teacher preferences						
Corr. R.C.	0.0188	0.0109	-0.0900	0.0088	0.0231	-0.0006
6. Vary window in which we estimate principal preferences: baseline is all applications						
W/in 2 weeks of hire	0.0222	0.0182	-0.0786	0.0103	0.0231	-0.0006
First half	0.0239	0.0204	-0.0793	0.0096	0.0234	-0.0013
Second half	0.0224	0.0161	-0.0787	0.0077	0.0242	-0.0005
7. Estimate principal preferences using rank order logit: baseline is binary logit						
All data	0.0201	0.0163	-0.0786	0.0064	0.0231	-0.0006
Active choices	0.0206	0.0191	-0.0793	0.0038	0.0234	-0.0013
Hire outcome only	0.0272	0.0233	-0.0787	0.0108	0.0242	-0.0005
8. Hold class sizes constant: baseline uses class size						
Constant class size	0.0223	0.0136	-0.0883	0.0540	0.0197	0.0660
9. Alternative value-added models						
Homogeneous	0.0210	0.0185	-0.0851	-0.0193	0.0053	-0.0271
Using school means	0.0156	0.0014	-0.0919	0.0319	-0.0056	0.0216
Using alternative FEs	0.0110	0.0054	-0.0805	-0.0176	0.0028	-0.0199
10. Add covariates to principal model						
	0.0206	0.0161	-0.0787	0.0094	0.0247	-0.0012

The table shows robustness checks for our main results. The columns show the difference in the predicted output for disadvantaged students minus advantaged students (a positive number indicates that disadvantaged students have better teachers). The "status quo" corresponds to the teacher-proposing DA with estimated teacher and principal preferences and restricted options based on timing. "All Options" expands teachers' choice sets to all positions, "Max VA" corresponds to ranking positions (or teachers) by predicted value-added, and "First Best" is the output-maximizing allocation. See footnote to Table 9 for descriptions of each of the robustness exercises.

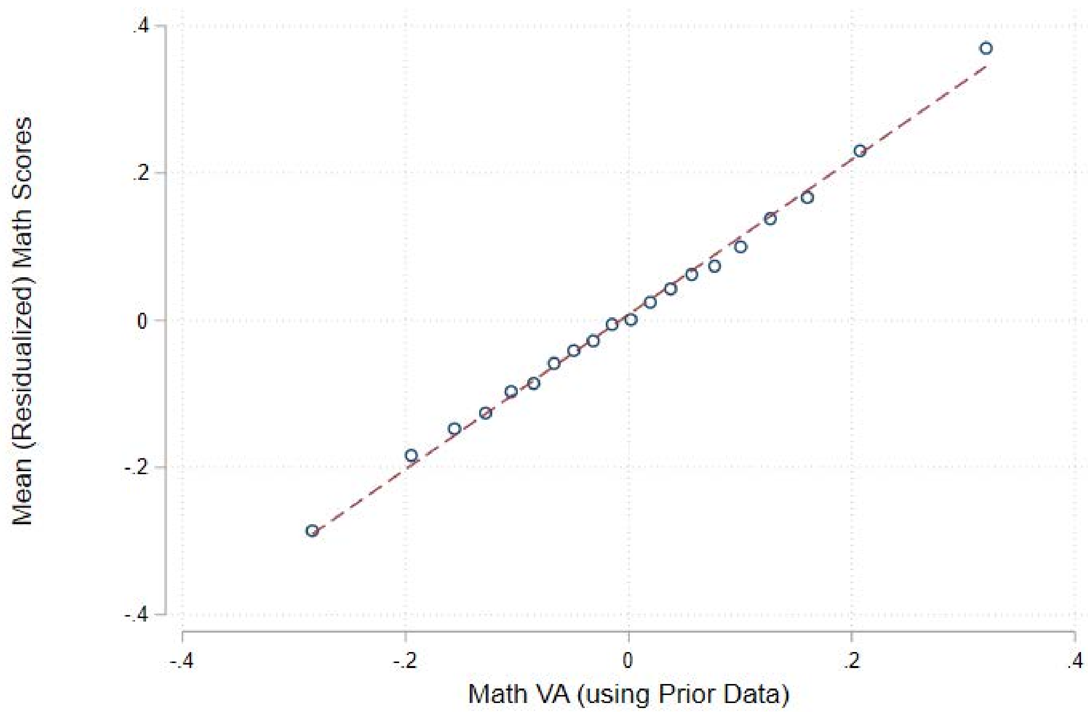
Table 9: Robustness: output relative to status quo

	All Options	Principal Max VA	Teach Max VA	Both Max VA	First Best
Baseline	0.0044	-0.0002	0.0216	0.0228	0.0309
1. Vary year: baseline is 2016					
2012	0.0025	0.0031	0.0238	0.0228	0.0328
2013	0.0028	0.0017	0.0234	0.0225	0.0313
2014	0.0015	0.0031	0.0217	0.0225	0.0312
2015	0.0031	-0.0091	0.0228	0.0243	0.0342
2017	0.0058	-0.0053	0.0244	0.0276	0.0363
2. Vary student type split: baseline is economic disadvantage					
Achievement	0.0052	-0.0003	0.0243	0.0277	0.0359
Race	0.0032	0.0002	0.0232	0.0275	0.0356
3. Vary choice set construction for teachers					
7 day buffer	0.0046	-0.0021	0.0215	0.0227	0.0309
First day choice sets only	0.0064	0.0011	0.0251	0.0263	0.0345
Drop single app. teachers	0.0032	-0.0026	0.0205	0.0221	0.0303
4. Vary teacher preference specification to use binary logit					
No REs or FEs	0.0034	-0.0011	0.0219	0.0232	0.0315
School FEs	0.0021	-0.0026	0.0232	0.0249	0.0330
School REs	0.0031	-0.0019	0.0215	0.0232	0.0312
Teacher FEs	0.0045	-0.0001	0.0213	0.0225	0.0307
Teacher REs	0.0037	-0.0004	0.0209	0.0225	0.0307
Teacher REs, School FEs	0.0019	-0.0091	0.0239	0.0256	0.0337
Teacher FEs, School FEs	0.0014	-0.0110	0.0248	0.0261	0.0344
5. Allow for correlated random coefficients in teacher preferences					
Corr. R.C.	0.0038	-0.0018	0.0230	0.0243	0.0324
6. Vary window in which we estimate principal preferences: baseline is all applications					
W/in 2 weeks of hire	0.0046	0.0001	0.0219	0.0231	0.0312
First half	0.0041	-0.0002	0.0215	0.0233	0.0314
Second half	0.0044	0.0001	0.0214	0.0236	0.0316
7. Estimate principal preferences using rank order logit: baseline is binary logit					
All data	0.0041	0.0004	0.0210	0.0233	0.0315
Active choices	0.0027	-0.0003	0.0194	0.0231	0.0313
Hire outcome only	0.0048	0.0006	0.0220	0.0241	0.0322
8. Hold class sizes constant: baseline uses class size					
Constant class size	-0.0005	-0.0029	0.0047	0.0045	0.0064
9. Alternative value-added models					
Homogeneous	0.0028	0.0012	0.0231	0.0283	0.0346
Using school means	-0.0034	-0.0041	0.0259	0.0395	0.0434
Using alternative FEs	0.0045	0.0039	0.0206	0.0238	0.0303
10. Add covariates to principal model					
	0.0041	-0.0005	0.0209	0.0227	0.0313

The table shows robustness checks for our main results. The columns correspond to the change in mean student achievement (in student standard deviation units) between the considered counterfactual and the estimated status quo. "All Options" expands teachers' choice sets to all positions, "Max VA" corresponds to ranking positions (or teachers) by predicted value-added, and "First Best" is the output-maximizing allocation. In the first section, we vary the year in which we implement our main exercise. In the second section, we show results where teacher-school match effects depend on different student observable characteristics. In the third section we vary the assumptions around teachers' choice sets or drop teachers who make single applications. In the fourth section, we vary the level of random or fixed effects in the teacher preference model, while in the fifth section we allow for correlated random coefficients on a constant, total value-added, and fraction of students who are economically disadvantaged. In the sixth section we vary principals' choice sets while in the seventh we vary how we treat an application's outcome in the principal preference model. In the eighth section we show results where preference estimation and counterfactual analysis use constant class sizes across all positions in the district.

In the ninth section, we vary the value-added model: (a) a homogeneous value-added model with constant effects across student types, (b) a model that uses school mean characteristics rather than school fixed effects, and (c) a model that uses teacher-year fixed effects, rather than teacher-class-student type fixed effects, in the first residualization step. In the tenth section, we add additional teacher observables from Table 2 to the principal model.

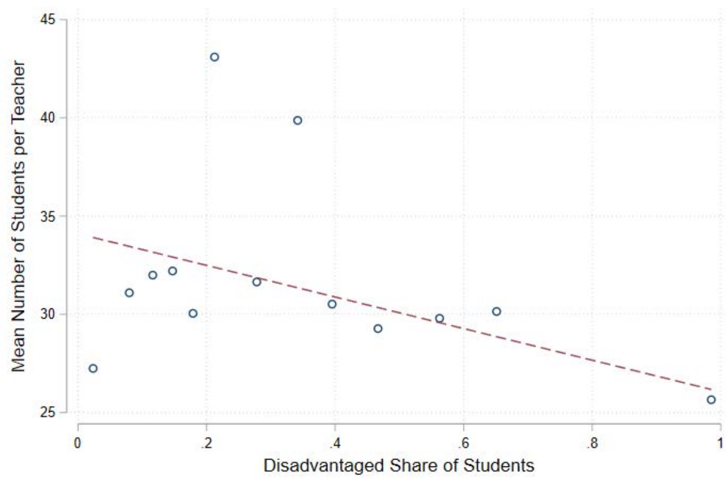
Figure 1: Math Value-Added Forecast Unbiasedness



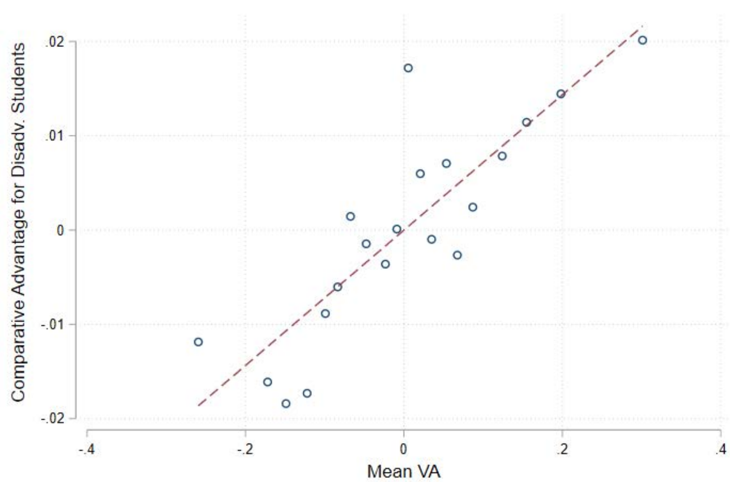
The figure is a binscatter, where an observation is a teacher-year and math value-added estimates are predictions using data from prior years. Units are student standard deviations. The y-axis is the mean student math test score, residualized by student demographics including lagged scores, school fixed effects, and teacher experience measures. The mean is taken over all students for a given teacher-year.

Figure 2: Features of classes and teachers

(a) Class size and fraction disadvantaged



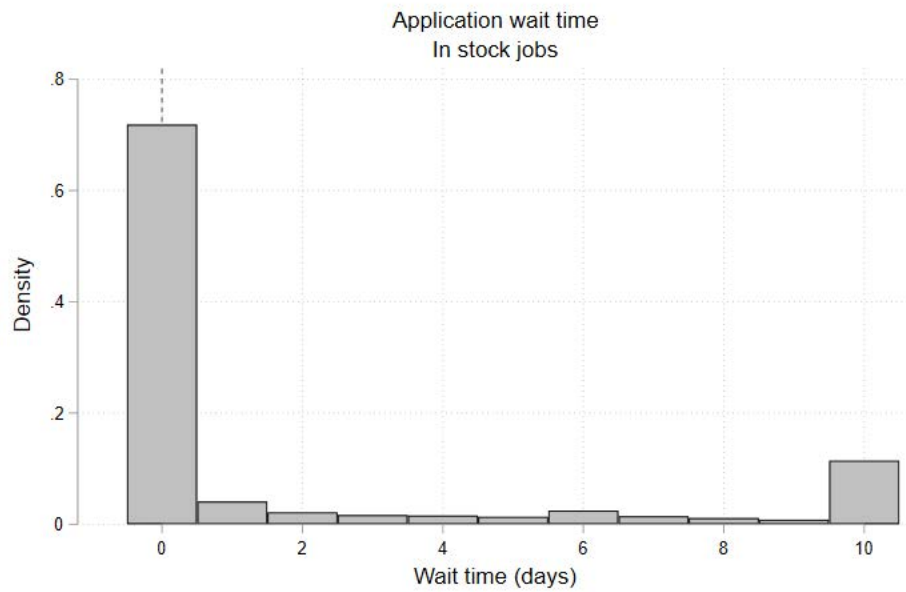
(b) Comparative advantage for disadvantaged students, and absolute advantage



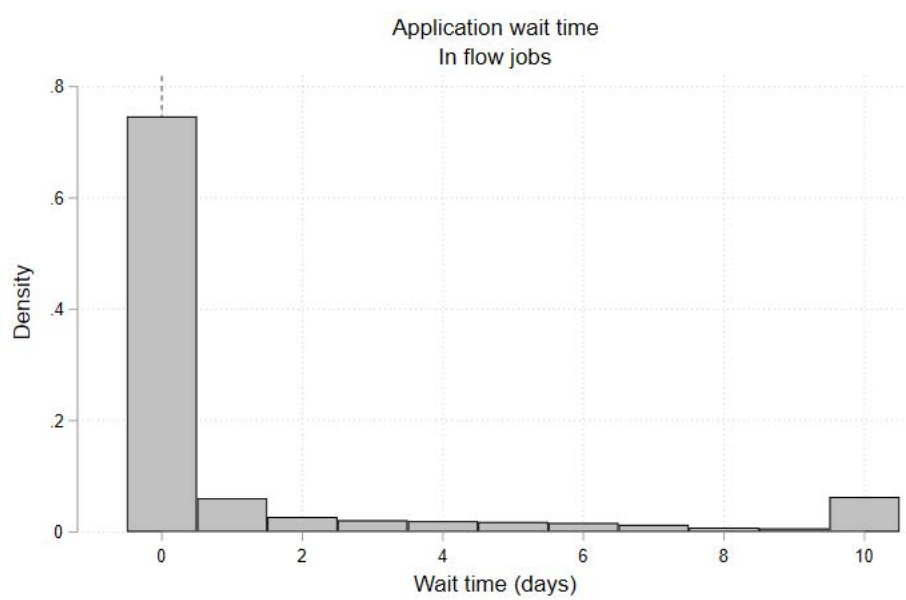
The figures show binscatters related to classroom characteristics and teacher characteristics. The top panel shows the relationship between a school's (mean) disadvantaged share of students and a school's (mean) number of students per teacher. The right-most point of the binscatter, with 100% of a school's students economically disdavantaged, accounts for 36% of the sample. The bottom panel shows the relationship between a teacher's absolute advantage (x-axis) and comparative advantage in teaching economically disadvantaged students (y-axis). For this figure, absolute advantage is the average value-added across students types (rather than the value-added at a representative school) to avoid mechanical correlations between absolute and comparative advantage.

Figure 3: Wait time to apply to vacancies

(a) Stock of vacancies



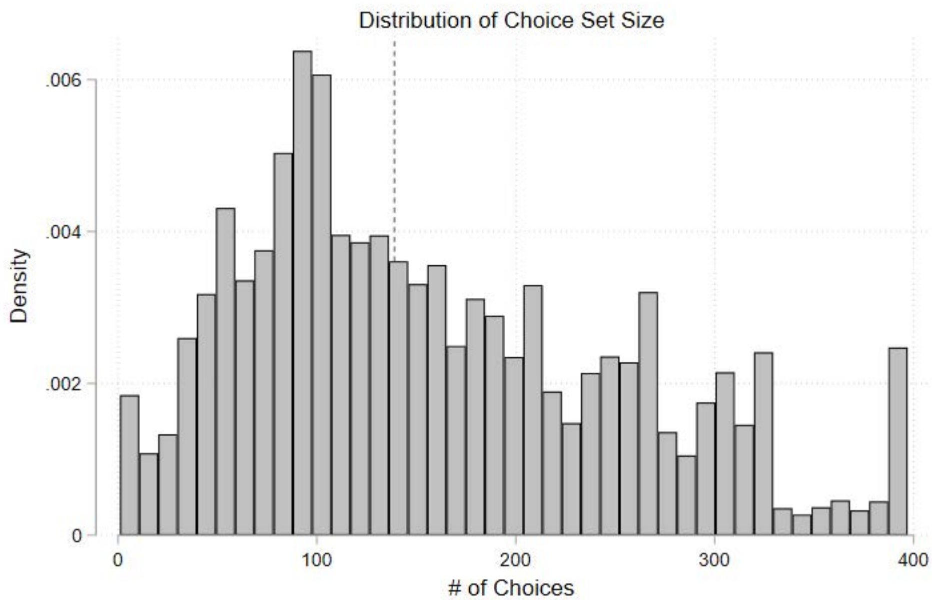
(b) Flow of vacancies



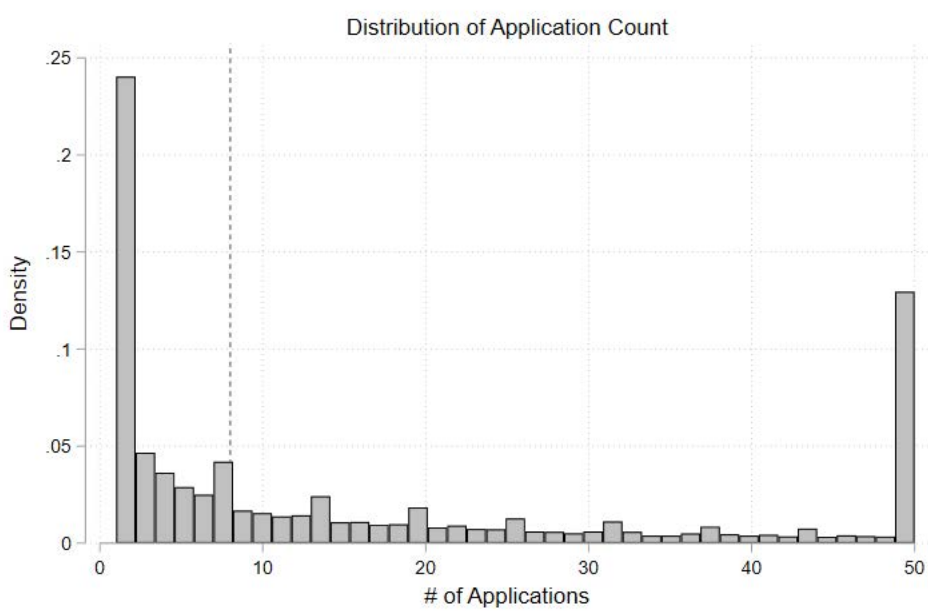
The figures show the wait time for applicants to apply to vacancies. In Panel A, we look at vacancies that were "in stock" (already posted) on the day the teacher first applied on the platform. We plot the "leave one out" wait time, where we omit one job the teacher applied to on the first day. In Panel B we look at the wait time to apply to vacancies that were posted after the teacher first applied on the platform. We measure wait time as the time from when the teacher first applied to another job (once the focal position is posted) until they apply to the posted job. We place vertical dashed lines at the median wait time.

Figure 4: Choice and application set sizes

(a) Choice sets

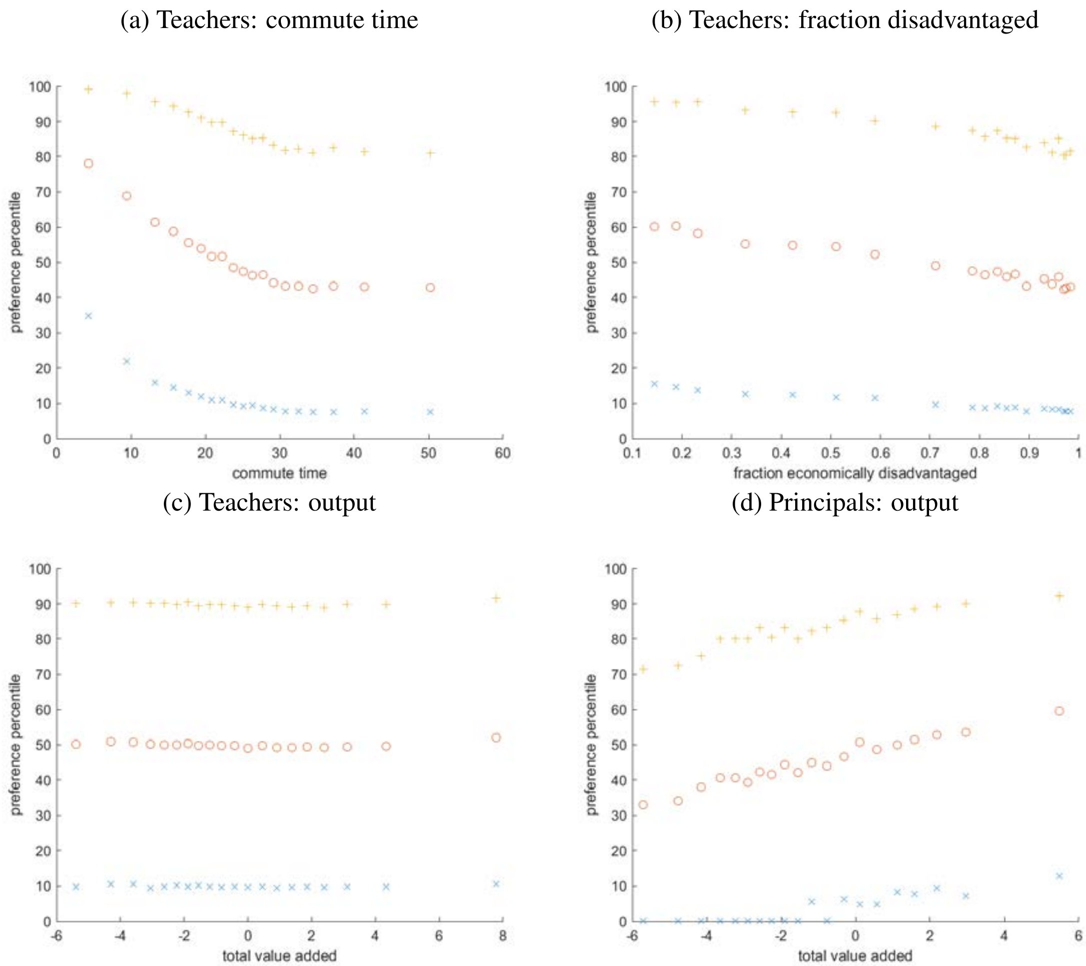


(b) Application sets



The figures are histograms of the number of positions a teacher has in her choice set (Panel A) and the number of positions a teacher applies to (Panel B). An observation is an applicant-year. Choice sets comprise the set of vacancies that are active while at some point between the teacher's first and last application in a given cycle.

Figure 5: Bivariate preference relationships



This figure shows binscatters of bivariate relationships between characteristics and preferences. In Panels (a)-(c), we show the bivariate relationship between characteristics in our teacher preference model and how teachers rank positions by estimating each teacher's ranking over positions and ordering positions from a teacher's most preferred (100) to least preferred (0). In Panel (d), we estimate show the bivariate relationship between characteristics in our principal model and principal rankings. We estimate each principal's ranking over teachers and order teachers from a principal's most preferred (100) to least preferred (0). The middle set of points (red circle) is the mean percentile, while the top (orange cross) and bottom (blue x) sets of points are the 10th and 90th percentiles, respectively.

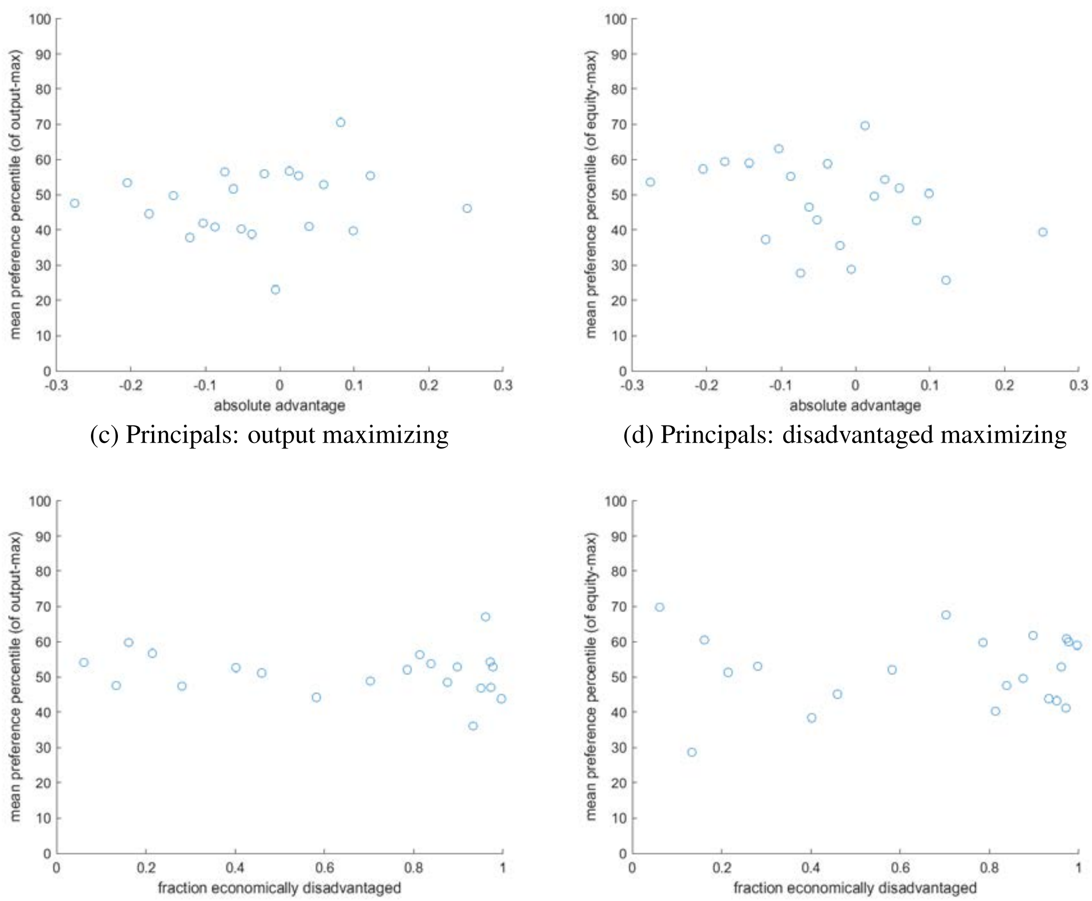
Figure 6: Rank of first-best for equity and efficiency in preferences

(a) Teachers: output maximizing

(b) Teachers: disadvantaged maximizing

(c) Principals: output maximizing

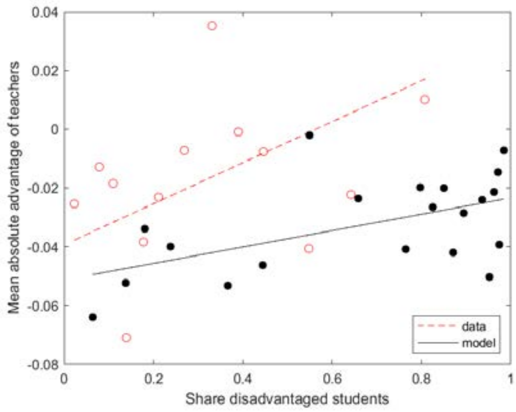
(d) Principals: disadvantaged maximizing



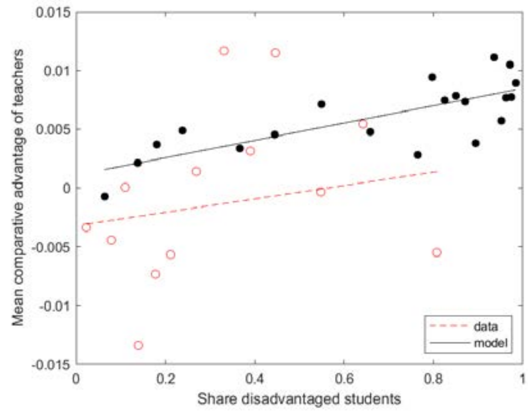
This figure shows how teachers and principals value the allocations associated with different first-best problems. In Panels (a) and (c) we calculate the output-maximizing first-best allocation. In Panel (a) we show the binscatter of how teachers rank the position they receive in the output-maximizing first-best allocation, by the teacher's absolute advantage. In Panel (b) we show the binscatter of how principals rank the teacher they receive in the output-maximizing first-best allocation, by the school's fraction of students that are economically disadvantaged. In Panels (b) and (d) we repeat the exercises for the first-best allocation that maximizes output for economically disadvantaged students. We construct teacher ranks over positions from our teacher preference model, ordering positions from a teacher's most preferred (100) to least preferred (0). We construct principal ranks over teachers from our principal model, ordering teachers from a principal's most preferred (100) to least preferred (0).

Figure 7: Model fit

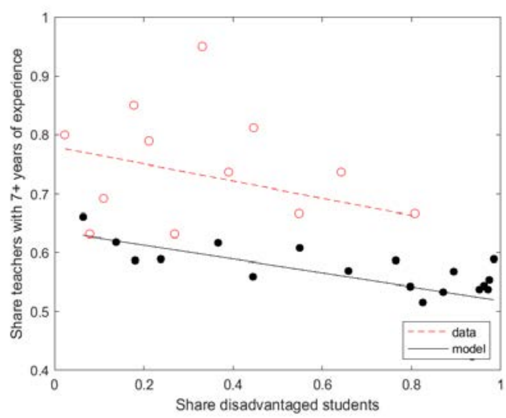
(a) Teacher absolute advantage



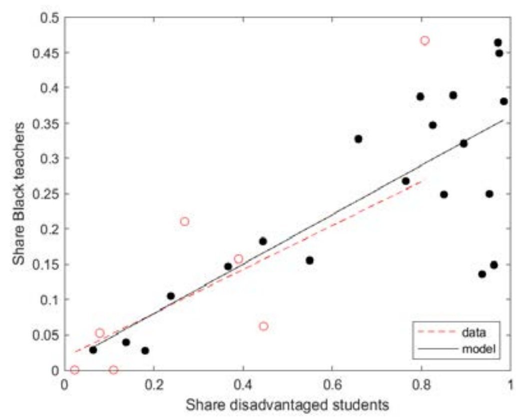
(b) Teacher comparative advantage



(c) Teachers with 7+ years of experience



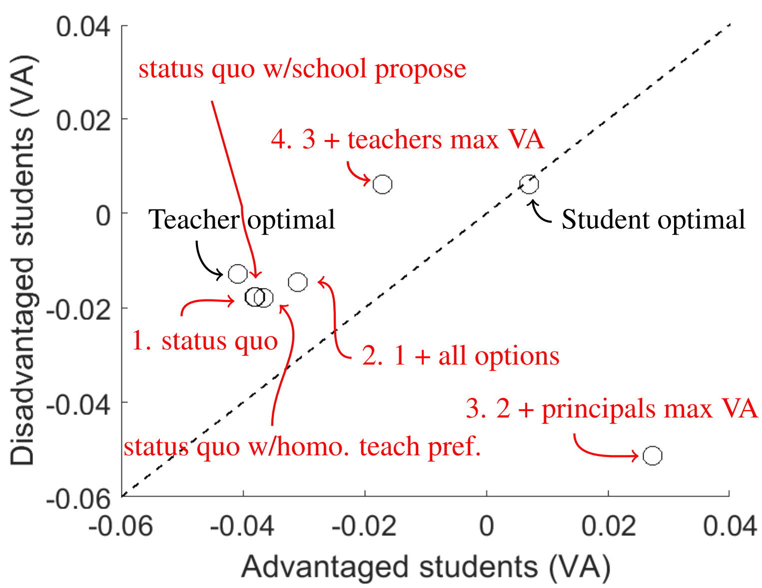
(d) Teachers that are Black



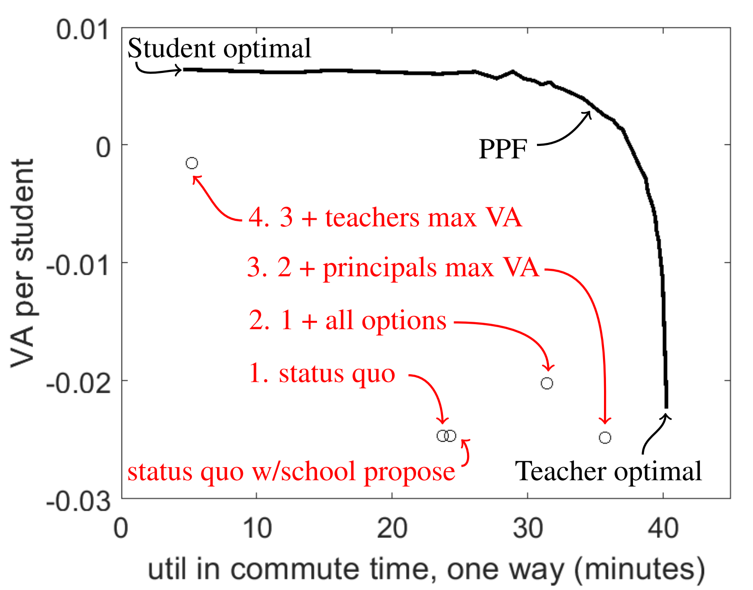
This figure compares the allocations implied by the model to the allocations we observe in the data. The data refers to all teachers in the district. The model refers to the teachers who apply in the transfer system for whom we have value-added scores. Positions are sorted on the x-axis by share of disadvantaged students.

Figure 8: Production possibilities frontier

(a) Distributional consequences



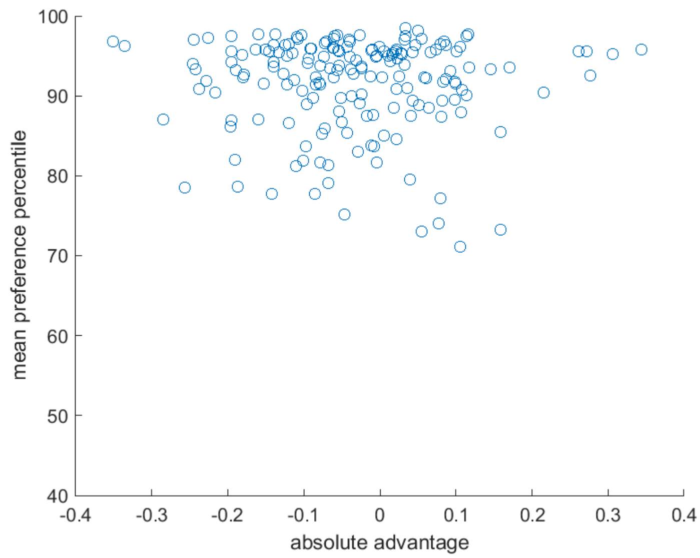
(b) Production possibilities frontier



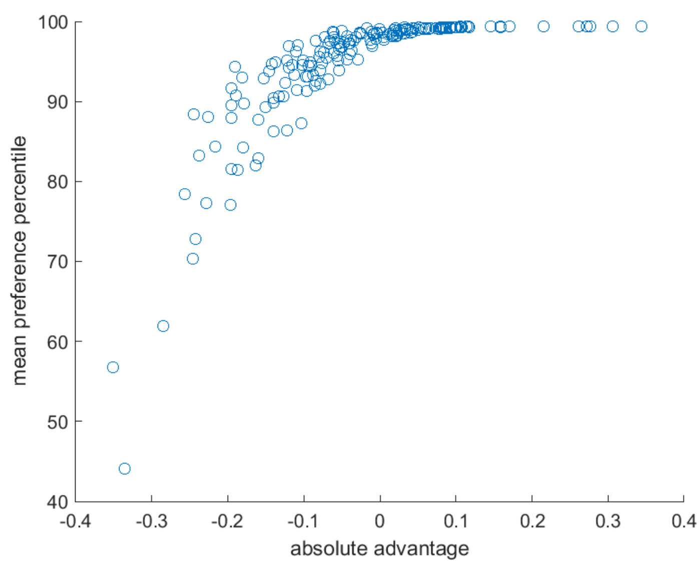
This figure simulates the trade-off between student achievement for economically advantaged and disadvantaged students (Panel A) and between teacher preferences and student achievement (Panel B). The "PPF" represents the solution to the social planner's problem from placing different relative weights on teacher preferences and student achievement. The student-optimal point maximizes student achievement and is when the planner only weights students. The teacher-optimal point maximizes teacher preferences and is when the planner only weights teachers. The status quo (point 1) uses teacher and principal estimated preferences, restricted choice sets, and solves for the teacher proposing stable allocation. The status quo with school proposing allocation is the same as the status quo except it is the school-proposing solution. Point 2 takes the status quo and gives teachers and principals all options. Point 3 takes point 2 and gives principals preferences to maximize value-added. Point 4 takes point 3 and also gives teachers preferences to maximize value-added. "Status quo w/homo. teach pref." shows the status quo where teacher preferences are estimated with a model that does not include random coefficients. The Figure plots averages over 200 simulations.

Figure 9: Teacher choice rankings

(a) Estimated principal preferences



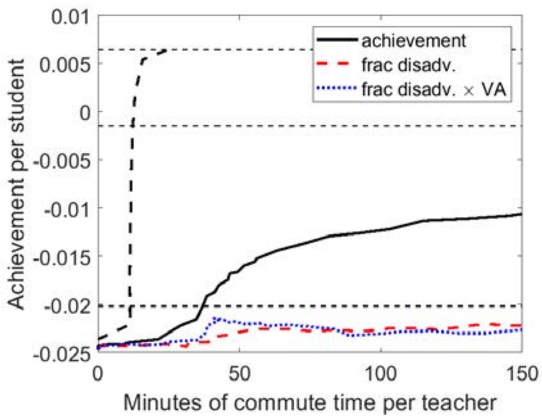
(b) Principals maximize value-added



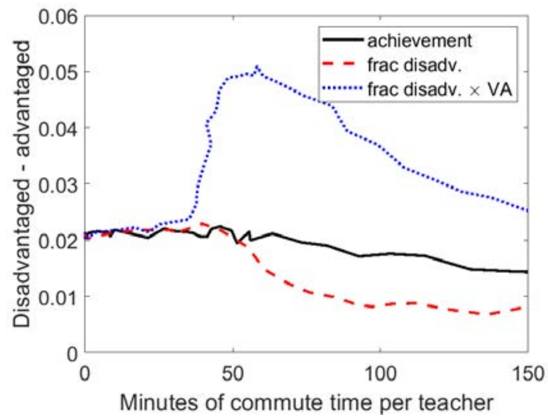
This figure presents the preference percentile of the position to which the teacher is assigned in two equilibria. We estimate a teacher's ranking of all positions and express it in percentiles, where 100 is the teacher's most preferred position. The top panel shows the status quo (point 1 in Figure 8). The bottom panel shows the same outcomes in point 3 in Figure 8: the status quo with the complete choice set, and principals maximize value-added. Teachers are ordered on the x-axis by their absolute advantage (predicted value-added at the district's representative school).

Figure 10: Bonus schemes

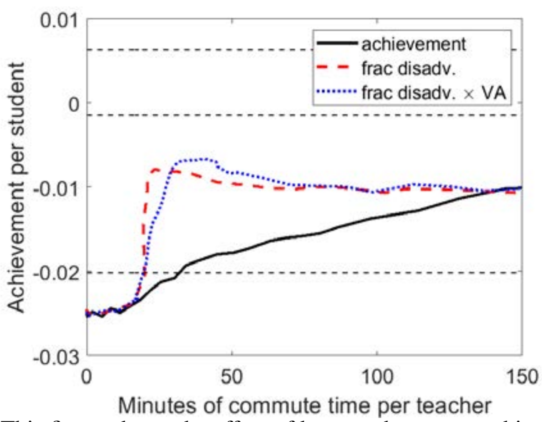
(a) Relative to the status quo, aggregate achievement



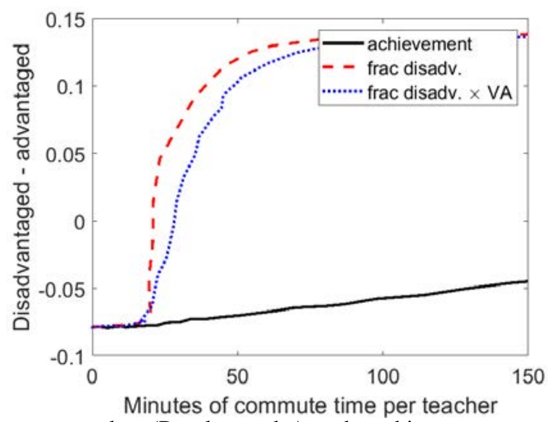
(b) Relative to the status quo, achievement gap



(c) With principal bonuses, aggregate achievement



(d) With principal bonuses, achievement gap



This figure shows the effect of bonus schemes on achievement per student (Panels a and c) or the achievement gap (Panels b and d). In all panels, the x-axis shows the cost of the policy per teacher, which we express in minutes of commute time per teacher. The y-axis shows the benefits in terms of achievement per student or the difference in achievement between disadvantaged and advantaged students. We consider three policies: subsidizing achievement directly, subsidizing the position based on the fraction of disadvantaged students in the position, and subsidizing the position based on fraction disadvantaged interacted with the teacher's absolute advantage. In the top two panels, we take as the baseline allocation the status quo, and the constant part of the bonus is chosen to make teachers weakly better off relative to this allocation. In the bottom two panels we replace estimated principal preferences with preferences that maximize output. The dashed line in the left panels is the cost of the first-best policy and represents movements along the PPF. The three horizontal dashed lines correspond to the output in the first-best (top), the output in the allocation where teachers and principals each maximize value-added and choice sets are complete (point 4 in Figure 8) (middle), and the output in the allocation where teachers and principals each maximize estimated preferences and choice sets are complete (point 2 in Figure 8 (bottom)).

A Data Appendix

A.1 Student-level data

We use student records from the NCERDC over the years of 2006-2007 through 2017-2018 to measure multi-dimensional teacher productivity in raising math test scores. This provides 8,177,312 student-year observations. We focus on math teachers in grades 4 through 8 to capture the majority of teachers with prior performance data who enter the applicant pool. We use third to seventh grade math and reading scores as lagged achievement. Test score data as well as student demographics such as ethnicity, gender, gifted designation, disability designation, whether the student is a migrant, whether the student is learning English, whether the student is economically disadvantaged, test accommodations, age, and grade come from the NCERDC master-build files. We use only data from standard end-of-grade exams. This leaves us with 5,322,896 student-year observations.

Beginning in the 2006-2007 school year, the state began recording course membership files linking students directly to courses and instructors. Prior to this change, teachers were linked to students through data on the proctors of the end-of-course exams. The new course membership files provide stronger teacher–subject-student links than the previous system, in which teachers were more frequently linked to the wrong subject (Harris and Sass, 2011).

With the course membership files, we still must determine which teacher is most responsible for teaching math. We use a tiered system. We use course codes (starting with "20") and course names (including text "math," "alg," "geom," and "calc") to do so. We also want to prioritize standard classes as opposed to temporary or supplemental instruction (course names including text such as "study," "special," "resource," "pullout," "remed," "enrich," "indiv," and "except"). We assign students to the teacher most likely to be the math teacher according to the following rules: (1) Students are assigned first to a high-certainty math teacher (the course code and title indicate a standard math class without mention of supplemental instruction). (2) Students with self-contained teachers are assigned to that teacher if there is no high-certainty math teacher present. (3) Students with course codes and course titles indicating math teachers but no self-contained teachers or high-certainty math teachers are assigned to those middle-certainty math teachers. (4) Students with a teacher of a course that either has a math code or a math course title but no other math course or self-contained teacher are assigned to those low-certainty math teachers. (5) Students with a science course code but no math course or self-contained courses are assigned to their science teachers to accommodate recent trends of math and science block scheduling. We exclude classes in which more than half the class requires special accommodations. Ultimately, our sample for constructing teacher value-added measures is composed of 5,159,337 student-year observations providing measures for 38,566 teachers.

A.2 Application and vacancy data

Our application and vacancy data cover the 2010-2019 cycles. We restrict our sample to applications and vacancies for on-cycle, standard elementary school positions. We show how these restrictions change the sample in Appendix Table A1.

We define on-cycle as positions that receive their first applications of a cycle between April 1 and August 15.

We select standard elementary school positions by filtering on the vacancy type (”instructional”) and the vacancy title. Seventy percent of posted vacancies are for instructional positions. We require that the position indicate elementary school grades by having at least one of the following text strings in the title: “k-”, “3rd”, “4th”, “5th”, “-5”, “-6”, “4-6”, or ”elem”. 39% of vacancies include at least one of these strings in the title.

We then exclude positions with specific subjects mentioned in the title or indications that the position is non-standard (“specialized”, “end of year”, “interim”, “assistant”, “virtual”, “resource”, “itinerant”, “exchange”, “extensions”, “immersion”, “academic support”, “temporary”, “continuous”, “early end”, “interventionist”, or “substitute”). With all of the restrictions above, our final sample consists of 20% of the full set of applications, 25% of the full set of applicants, and 7% of the full set of vacancies.

We code the application’s outcome into whether the candidate is hired (“Accepted-Pending Licensure”, “Hired”, “Hiring Request in Process”, “Offer Accepted”), declines an offer (“Offer Declined”), offered an interview (“Completed BEI Interview”, “Contact for Interview”, “Interview Scheduled”, “Invited to Complete Virtual Interview”, “Invited to Interview”, “Recommended for Interview (By Request)”), or given a positive rating (“1st Choice”, “2nd Choice”, “Highly Recommend for Interview”, “Recommend”, “Recommend for Interview”, “Recommendation Accepted”, “Strong Candidate”). These categories are encodings of a single variable, so they are mutually exclusive (i.e., if a candidate is hired, the prior outcome may be overwritten). For robustness analysis, we also split up the remaining applications into middle ratings (“Attended Info Session/Class”, “Hold for Later Consideration”, “Invited to Info Session/Class”, “Possible recommend for interview”, “Recommend with Hesitation”), negative ratings (“Failed Job Questionnaire”, “Incomplete Application”, “Ineligible Selection”, “Not Good Fit”, “Not Qualified”, “Pool - Ineligible”, “SS - INELIGIBLE”, “Screened - Not Selected”), withdrawals (“Candidate Withdrew Interest”), or no evaluation (“Eligible Selection”, “New”, “Pool - Eligible”, “Pool Candidate”).

A.3 Matching across datasets

For this project the North Carolina Education Research Data Center (NCERDC) combined records held there on teacher work histories, school characteristics, and student achievement with data provided by a large urban school district containing further personnel files, open positions within the school district, and applications for those positions. They performed an interactive fuzzy match using names and birth year. For teachers who had a sufficiently good match (that is, a unique name-birth-year combination), we have a de-identified ID that allows us to connect their platform data to their staffing records and students' achievement.

The NCERDC reports that of the 74,395 applicants to positions, 29,008 are matched to NCERDC records. Many of these applicants never teach in the state and thus would not be expected to match. Of the 26,983 employees listed within the district, 20,966 are matched to NCERDC records. However, the match rate is much better among personnel who teach tested subjects. Of the 13,982 teachers with EVAAS scores in the district, 13,865 are matched to the NCERDC data.

A.4 Sample characteristics

Returning to Appendix Table A1, we see how the sample's characteristics varies with sample restrictions. The "Elementary Sample" restricts to on-cycle elementary school instructional positions without specialization, the "Value-Added Sample" further restricts to teachers with value-added forecasts based on prior years, and the "2015 Sample" further restricts to the 2015 application cycle. We use the "Elementary Sample" for estimating principal preferences, the "Value-Added Sample" for estimating teacher preferences, and the "2015 Sample" for estimating counterfactual allocations.

We see a few expected patterns based on the sample restrictions. For the last two columns, we require teachers to have value-added forecasts based on data from prior years. This restrictions leads us to a more experienced sample of teachers. These teachers are more likely both to already be in the district and to transfer to a new school (from a prior school or from out of district). We also see these teachers have lower application rates, perhaps because many already have in-district placements. We see little change in the teacher sample's mean value-added (by student type or at a representative school) or choice set size. The mean characteristics in the positions sample also change minimally with the sample restrictions.

B Omitted details on value-added model: assumptions, results, and validation

B.1 Formal statement of assumptions for value-added model

Here we formally state the assumptions that were informally discussed in Section 3.

Assumption 1 (Exogeneity and stationarity of classroom and student-level shocks)**.** *Classroom-student-type shocks (θ_{cmt}) are independent across classrooms and independent from teachers and schools. Classroom-student-type shocks follow a stationary process:*

$$\mathbb{E}\left[\theta_{c0t}|t\right] = \mathbb{E}\left[\theta_{c1t}|t\right] = 0 \tag{A1}$$

$$Var(\theta_{c0t}) = \sigma^2_{\theta_0},\ Var(\theta_{c1t}) = \sigma^2_{\theta_1},\ Cov(\theta_{c0t}, \theta_{c1t}) = \sigma_{\theta_0\theta_1} \tag{A2}$$

for all t.

Student-level idiosyncratic variation is independent across students and independent from teachers and schools. Student-level shocks follow a stationary process depending on the student's type:

$$\mathbb{E}[\tilde{\varepsilon}_{it}|t] = 0 \tag{A3}$$

$$Var(\tilde{\varepsilon}_{it}) = \sigma^2_{\varepsilon m}\ \textit{for}\ m = 0, 1 \tag{A4}$$

for all t.

Assumption 2 (Joint stationarity of teacher effects)**.** *The non-experience part of teacher value-added for each student type follows a stationary process that does not depend on the teacher's school. The covariances between the teacher's value-added across student types depend only on the number of years elapsed:*

$$\mathbb{E}[\mu_{j0t}|t] = \mathbb{E}[\mu_{j1s}|t] = 0 \tag{A5}$$

$$Var(\mu_{j0t}) = \sigma^2_{\mu_0},\ Var(\mu_{j1t}) = \sigma^2_{\mu_1},\ Cov(\mu_{j0t}, \mu_{j1t}) = \sigma_{\mu_0\mu_1} \tag{A6}$$

$$Cov(\mu_{j0t}, \mu_{j0,t+s}) = \sigma_{\mu_0 s},\ Cov(\mu_{j1t}, \mu_{j1,t+s}) = \sigma_{\mu_1 s} \tag{A7}$$

$$Cov(\mu_{j0t}, \mu_{j1,t+s}) = \sigma_{\mu_0\mu_1 s} \tag{A8}$$

for all t.

Assumption 3 (Independence of drift and school effects)**.** *Let $\bar{\mu}_{jm}$ be teacher j's mean value-added for student type m. Let k be j's assigned school in year t. Then:*

$$(\mu_{jmt} - \bar{\mu}_{jm}) \perp \mu_k\ \textit{for}\ m = 0, 1. \tag{A9}$$

B.2 Additional details on estimation

In the first step, we estimate β_l by regressing test scores (standardized to have mean 0 and standard deviation 1 in each grade-year) on a set of student characteristics (X_{it}) and classroom-student-type fixed effects:

$$A^*_{it} = \beta_s X_{it} + \lambda_{cmt} + \upsilon_{it}. \tag{A10}$$

For characteristics, we include ethnicity, gender, gifted designation, disability designation, whether the student is a migrant, whether the student is learning English, whether the student is economically disadvantaged, test accommodations, age, and grade-specific cubic polynomials in lagged math and lagged reading scores. We subtract the estimated effects of the student characteristics to form the first set of residuals, $\hat{\nu}_{it}$:[25]

$$\hat{\nu}_{it} = A^*_{it} - \hat{\beta}_s X_{it}. \tag{A11}$$

These student-level residuals include teacher, school, and classroom components, as well as idiosyncratic student-level variation.

In the second step, we project the residuals onto teacher fixed effects, school fixed effects, and the teacher experience return function. Following the literature, we specify the experience return function as separate returns for every level of experience up to 6 years, and then a single category of experience of at least 7 years:

$$\hat{\nu}_{it} = \sum_{e=1}^{6} \alpha^e \mathbb{1}\{Z_{jt} = e\} + \alpha^7 \mathbb{1}\{Z_{jt} \geq 7\} + \mu_{jm} + \mu_k + \mu_t + \varepsilon_{it}, \tag{A12}$$

where $\varepsilon_{it} = (\mu_{jmt} - \mu_{jm}) + \theta_{cmt} + \tilde{\varepsilon}_{it}$. We then form a second set of student-level residuals by subtracting off the estimated school and experience effects:

$$A_{it} = \hat{\nu}_{it} - \left(\sum_{e=1}^{6} \hat{\alpha}^e \mathbb{1}\{Z_{jt} = e\} + \hat{\alpha}^7 \mathbb{1}\{Z_{jt} \geq 7\} + \hat{\mu}_k + \hat{\mu}_t \right). \tag{A13}$$

We aggregate these student-level residuals into teacher-year mean residuals for each student type: $\bar{A}_{jmt}$. Let $\mathbf{A}_j^{-t}$ be a vector of mean residuals for each student type-year that j teaches in the data, prior to year t.

In the final step, we follow Delgado (2021) and form our estimate of teacher j's value-added (net of experience effects) in year t for type m as the best linear predictor based on the prior data in our sample:

$$\hat{\mu}_{jt} \equiv \mathbb{E}^* \left[\mu_{jt} | \mathbf{A}_j^{-t} \right] = \psi' \mathbf{A}_j^{-t}, \tag{A14}$$

where μ_{jt} is a $(2x1)$ vector for the teacher's output across the two student types and ψ is a $2(t-1)x(t-1)$ matrix of reliability weights where $t-1$ is the number of years of prior data. These weights minimize the mean squared error between the estimate of the teacher's value-added and our forecast based on prior data:

$$\hat{\psi}' = argmin \sum_j (\bar{A}_{jt} - \psi' \mathbf{A}_j^{-t})'(\bar{A}_{jt} - \psi' \mathbf{A}_j^{-t}). \tag{A15}$$

We estimate ψ following Delgado (2021). Here we describe how we estimate the structural parameters: $\sigma_{\varepsilon 0}, \sigma_{\varepsilon 1}, \sigma_{\theta 0}, \sigma_{\theta 1}, cov(\theta_{c0t}, \theta_{c1t}), \sigma_{\mu 0}, \sigma_{\mu 1}, cov(\mu_{j0t}, \mu_{j1t}), cov(\mu_{j0t}, \mu_{j0s}), cov(\mu_{j1t}, \mu_{j1s}), cov(\mu_{j0t}, \mu_{j1s})$.

- $\hat{\sigma}_{\varepsilon m} = \frac{1}{N_c} \sum_{c=1}^{N_c} \frac{1}{n_{cm}-1} \sum_{n=1}^{n_{cm}} (\hat{v}_{it} - \frac{1}{n_{cm}} \sum_{n=1}^{n_{cm}} \hat{v}_{it})$

[25] Here we deviate from the standard notation, by introducing $\hat{\nu}_{it}$. Our procedure has two residualization steps because we include classroom-student type fixed effects in the first step, which would subsume the teacher and school fixed effects. We thus decompose student residuals into teacher and school components in a second step.

- $\hat{\sigma}_{\theta m} = Var(\bar{A}_{jmtc}) - \hat{\sigma}_{\mu m} - \frac{1}{N_{cm}} \sum_{i=1}^{N_{cm}} \frac{\hat{\sigma}_{\varepsilon m}}{n_{cm}}$
- $\hat{cov}(\theta_{c0t}, \theta_{c1t}) = cov(\bar{A}_{j0tc}, \bar{A}_{j1tc}) - \hat{cov}(\mu_{j0t}, \mu_{j1t})$
- $\hat{\sigma}_{\mu m} = \sqrt{cov(\bar{A}_{jmtc}, \bar{A}_{jmtc'})}$, where $c \neq c'$
- $\hat{cov}(\mu_{j0t}, \mu_{j1t}) = cov(\bar{A}_{j0tc}, \bar{A}_{j1tc'})$, where $c \neq c'$
- $\hat{cov}(\mu_{j0t}, \mu_{j0s}) = cov(\bar{A}_{j0t}, \bar{A}_{j0s})$
- $\hat{cov}(\mu_{j1t}, \mu_{j1s}) = cov(\bar{A}_{j1t}, \bar{A}_{j1s})$
- $\hat{cov}(\mu_{j0t}, \mu_{j1s}) = cov(\bar{A}_{j0t}, \bar{A}_{j1s})$

where N_c is the number of classes, N_{cm} is the number of classes times student types, and n_{cm} is the number of students in class c of type m,

Our estimate of teacher j's composite value-added at school k in year t is:

$$\widehat{VA}_{jkt} = p_{k0t}\hat{\mu}_{j0t} + p_{k1t}\hat{\mu}_{j1t} + f(Z_{jt}; \hat{\alpha}). \tag{A16}$$

Variation in the data: We now discuss the variation in the data that pins down key parameters. The coefficient on student characteristics uses how test scores vary with within-classroom-student type variation in student characteristics.[26] The school effects use the change in (student) output when teachers switch schools, beyond what would be predicted by drift and by the change in student type composition. Heuristically, if teachers' output regularly increases when teachers transfer to a certain school, then we would estimate a high school effect. The teacher mean effects for each student type are pinned down by relative increases in students' (residualized) test scores across different teachers. We are able to rank teachers both within and across schools, provided teachers and schools are in a set connected by transfers so that we can identify the school effects.

Finally, we identify the parameters of the teacher value-added distribution and the drift process based on the stationarity assumptions and the observations of teachers across years, classrooms, and student types. As an example, the variance of the teacher effects for student type m is identified by the covariance between a teacher's mean student residuals for student type m in two different classrooms in the same year.[27] With our assumptions that classroom and student shocks are uncorrelated across classrooms, the only reason a teacher's students would have similar (residualized) outcomes is the teacher's value-added.

[26] Because we include classroom-student-type fixed effects, our model allows for an arbitrary correlation between students' characteristics and the quality of their assigned teachers. Allowing such correlation is important in a context where teachers have some control over where they work.

[27] In our setting many elementary school teachers have students from multiple classes. The prevalence of multiple classrooms is increasing over time (Appendix Table A11).

The first key parameter estimate is the significant dispersion in value-added for both student types of about 0.24σ. The second key parameter estimate is the strong correlation of 0.86 between the teacher's value added with the two types of students (Appendix Table A4). We find large returns to experience in the first year, and then a profile that flattens out after about four years of experience (Appendix Table A12). Appendix Figure A21 plots the drift parameters.

B.3 Alternative estimators

In our analysis, we explore the robustness of our results to elements of our value-added model. We focus on three variations from our baseline model.

Homogeneous: We estimate a model where teachers have a homogeneous effect on students' test scores and classroom shocks are not specific to student type:

$$\begin{aligned} \mu_{j0t} &= \mu_{j1t} = \mu_{jt} \\ \theta_{c0t} &= \theta_{c1t} = \theta ct \end{aligned} \tag{A17}$$

Using school means: In our baseline model, we include school fixed effects: μ_k. For robustness, instead of including μ_k in Equation A12, we include school-level means for all of the variables in X_{it}. Note that this will not deliver identical estimates because we do not include school-level means of the teacher fixed effects.

Using alternative fixed effects: In our baseline model, we include teacher-class-student type fixed effects (λ_{cmt}) in our first residualization step (Equation A13). For robustness, we include teacher-year fixed effects (λ_{jt}).

B.4 Testing for comparative advantage

Our measures forecast teachers' future value-added without bias. Our high estimated correlation between a teacher's effectiveness with the two student types raises the question of whether our estimates of comparative advantage simply reflect statistical noise. We perform three exercises to test our multi-dimensional value-added model versus a single-dimensional model.

First, we estimate standard errors and confidence intervals for the structural parameters in our production model. The estimated correlation in teacher value-added across student types is 0.86. We can, however, decisively reject a correlation of 1 as the bootstrap standard error is 0.035, with a 95% confidence interval of $(0.73, 0.87)$ (Appendix Table A4).

Second, we perform a likelihood-ratio test comparing our model with a model with one-dimensional teacher value-added. We take the mean residuals at the level of the teacher-classroom-student type, $\bar{A}_{jcmt}$, and collect a teacher's mean residuals across classrooms and student types, which come from a normal distribution:

$$\begin{pmatrix} \bar{A}_{jc1t} \\ \bar{A}_{jc'2t} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\mu_1} + \sigma^2_{\theta_1} + \frac{\sigma^2_{\varepsilon_1}}{N_{jc0t}} & \sigma_{\mu_1\mu_2} \\ \sigma_{\mu_1\mu_2} & \sigma^2_{\mu_2} + \sigma^2_{\theta_2} + \frac{\sigma^2_{\varepsilon_2}}{N_{jc2t}} \end{pmatrix} \right). \tag{A18}$$

We compare the likelihoods across our baseline model and an alternate model of homogeneous value-added where $\sigma^2_{\mu_1} = \sigma^2_{\mu_2}$, $\sigma^2_{\theta_1} = \sigma^2_{\theta_2}$, $\sigma^2_{\varepsilon_1} = \sigma^2_{\varepsilon_2}$, and $\sigma_{\mu_1\mu_2} = 0$. Our likelihood-ratio test has 4 degrees of freedom, and we reject the homogeneous value-added model in favor of the heterogeneous model, with a test statistic of 610, so the p-value is arbitrarily small ($p < 0.0001$).[28]

Third, we fix a teacher's type according to whether she is above or below the median in comparative advantage in teaching economically disadvantaged students in pre-transfer schools. We then test whether changes in the share of economically disadvantaged students differentially predict changes in student test score residuals ($\hat{\nu}_{it}$ from equation A13) in post-transfer schools by teacher-type. The logic of the test is as follows. Under a homogeneous value-added model, changes in the share of economically disadvantaged students should have no bearing on changes in teacher productivity across schools. If our estimated comparative advantage is meaningful, however, then as the share of disadvantaged students rises, teachers with a comparative advantage in teaching disadvantaged students should see gains in average productivity relative to teachers with a comparative advantage in teaching economically advantaged students. Accordingly, we regress across-transfer changes in teacher-by-school average student residuals on across-transfer changes in the share of disadvantaged students interacted with teachers' type. The results appear in Appendix Table A13. For teachers with a comparative advantage in teaching advantaged students in pre-transfer schools, productivity falls as the share of disadvantaged students rises (p-value=0.043). In contrast, for teachers with a comparative advantage in teaching disadvantaged students, productivity rises as the share of disadvantaged students rises (p-value=0.014). These findings indicates that comparative advantage is persistent across settings and predictive of match-specific productivity.

C Within-school assignments

Our analysis focuses on the allocation of teachers across schools in a district. Another margin of allocation could be within-school assignment of teachers based on class size or composition. Ignoring this margin could affect our results in two ways. First, we could understate the potential allocation gains (or even focus on the less important margin). In Table 3 we show that the gains to within-school reallocation are much smaller than the gains from reallocation across schools.

Second, if within-school position characteristics are endogenous, our preference model might

[28] We restrict the sample to one randomly-chosen vector of mean residuals per teacher so that the observations in our likelihood are independent. We also find a similar test statistic when we use mean residuals, $\bar{A}_{jcmt}$, from a model where the fixed effects in the residualizing steps are not separated by student type.

be misspecified. For example, suppose that an experienced teacher can negotiate for the Honors class at a school but the inexperienced teacher cannot. We assess this possibility in two ways.

C.1 Persistence of classroom characteristics

If within-school assignment characteristics were endogenous and a function of the teacher's type, we would expect persistence in these characteristics over time. In Appendix Tables A14 and A15, we show the autocorrelations in number of students taught by a teacher and the fraction of students that are economically disadvantaged. In each table's top panel, we show the school-level autocorrelation. We find that differences across schools – which we leverage in our analysis – are fairly persistent. In each table's bottom panel, we show the teacher-level autocorrelation where we residualize by school-year fixed effects to isolate the within-school deviation. These within-school differences across teachers – which we do not leverage in our analysis – are not persistent at all.

C.2 Do teachers bargain over student assignment on the job market?

Second, we examine how students are assigned to teachers within and across schools. This question is of particular interest since we would like to know whether teachers bargain with principals over their student assignments. Are sought-after teachers assigned "preferable" class compositions? The primary teacher characteristic we use is *experience*, which principals value and is reliably measured in our data. We first explore the relationship visually. Student attributes have a linear relationship with log(experience), so we estimate models in which the outcome variables are student attributes and the primary explanatory variable is a teacher's log(experience). In regressions, standard errors are clustered by teacher and by year.

Without controlling for school setting, there is a strong relationship between experience and student attributes (see Appendix Table A16). More experienced teachers are assigned higher-scoring students, fewer economically disdavantaged students, more students designated as gifted, and fewer Black students.

Much sorting takes place across schools—as teachers gain experience, they sort to more suburban schools where students are less economically disadvantaged and higher achieving. In the basic cross section, we find that a 100 percent increase in experience reduces poverty shares by 0.037 percentage points (significant at the 0.001 level). When we control for year and school fixed effects, the coefficient on (log) experience falls by over 80 percent to 0.006 (significant at the 0.001 level). We examine the gradient among newly hired teachers. This group is of particular relevance because applicants (as opposed to teachers not applying to new jobs) are the teachers we consider in our counterfactual exercises. When looking at this group, we find no significant relationship between experience and disadvantaged-student assignment, conditional on school-year fixed effects. This suggests that principals do not sort students to teachers based on experience within a school, and indicates that bargaining over student characteristics is unlikely.

The pattern of sorting Black students to teachers is quite similar. We find that doubling teacher experience reduces the Black share of a teacher's class by 0.033 percentage points (significant at the 0.001 level). When looking within a school, the experience gradient falls by 97 percent—the sorting of Black students to teachers is almost exclusively across schools. When we examine the relationship among new hires, the relationship is even smaller and statistically insignificant. The gradient between student test scores and teacher experience attenuates by 90 percent when accounting for school-year fixed effects. There still is a small, systematic difference in test scores which appears to arise from hiring more experienced teachers to serve in gifted-and-talented classrooms. We see very experienced teachers assigned somewhat less desirable class assignments than would be predicted by the rest of the support. It may be that schools encourage older teachers to leave by giving them more difficult class compositions.

In summary, among new teachers, experience does not significantly predict assignment to disadvantaged students or Black students within schools. There is a small experience gradient for higher achieving students among new teachers. It seems teachers of gifted-and-talented classrooms tend to be senior.

D Heterogeneity in application rate gap between Title I and non-Title schools

To showcase unobservable preference heterogeneity, we focus on teacher preferences over a binary characteristic: whether the school has Title I designation. Appendix Table A17 shows that on average teachers are less likely to apply to Title I schools. The application rate to non-Title I schools is almost 16% and to Title I schools is about 14%, and leaving an application gap of close to 2 percentage points (or 10%).

The second and third columns of Appendix Table A17 show why we are able to estimate heterogeneity precisely: the median number of applications choices that each teacher makes is over 65 for both Title I positions and non-Title I positions. Thus, teachers' application sets have the potential to include many or few Title I positions.

Appendix Figure A22 shows that the mean gap in application rates across school types masks substantial heterogeneity. For each teacher, we calculate the gap in application rates (for positions in the teacher's choice set) between Title I and non-Title I schools, and then we plot the distribution of the gap. Visually, the distribution almost appears centered on zero (the median is 0.003). But there is substantial dispersion: the standard deviation of the raw gaps is 0.134.

Naturally, the standard deviation of the raw gap overstates the extent of dispersion because it incorporates noise. We now describe and implement a simple minimum distance estimator for the true standard deviation of the applicant gap. For each teacher j we observe a_{j1} applications sent to a Title I school and c_{j1} is the number of Title I vacancies in the teacher's choice set. We can then estimate

$$\hat{p}_{j1} = \frac{a_{j1}}{c_{j1}}$$

or teacher j's application probability to a Title I school.

Using the natural notation for a "not-Title I" school, we also have:

$$\hat{p}_{j0} = \frac{a_{j0}}{c_{j0}}$$

We can then compute the "gap", or Title I penalty, as

$$\hat{g}_j = \hat{p}_{j1} - \hat{p}_{j0}.$$

We are then interested in the distribution of these gaps – e.g., the standard deviation (sd) of g_j. Naturally, taking $sd(\hat{g}_j)$ will result in an over-estimate of the amount of dispersion.

We fit the following model.

$$p_{j0} = \hat{\bar{p}}_0$$

$$p_{j1} = N(\hat{\bar{p}}_1, \sigma)$$

where $\hat{\bar{p}}_{\cdot}$ are the population average application probabilities and σ is a parameter to estimate. We estimate σ by simulated method of moments where the moment to match is $sd(\hat{g}_j)$ and we simulate data from the model embedded in the previous two equation using the observed $\{c_{j0}, c_{j1}\}$.

We find that the estimated standard deviation is 0.114, so the visual depiction of noise is in line with the underlying truth. If we take the minimum distance estimate at face value, while on average teachers have higher application probabilities to non-Title I schools, about 44% of teachers have higher application probabilities to Title I instead. Hence, this suggests that even though on average teachers prefer non-Title I schools, there is a substantial amount of heterogeneity in the applicant pool. Depending on how such preference heterogeneity maps into the existing allocation of teachers to schools, policies that make Title I schools more attractive could have small or large effects on teachers' application rates.

E Selection into the transfer market

What explains the differences in student gains between Table 3 and the results depicted in Figure 8a? Here, we compare the teacher transfer market to the broader sample of teachers and positions. We first consider the representation of schools in the transfer market. Unsurprisingly, we see significant over-representation for positions in schools with high proportions of economically disadvantaged students. Appendix Table A18 shows that a 10 percentage point increase in the share of economically disadvantaged students is associated with 0.15 more positions posted. Because the overrepresented type of school is already the more common one (more than half of the students in our district are economically disadvantaged) this means that gains from sorting on comparative advantage are going to be understated in our transfer sample.

The pattern is less pronounced for the number of students a teacher instructs. Though the point estimate implies that an additional 10 students per teacher lowers the number of positions a school posts by 0.2, this relationship is largely driven by outliers, as shown in Appendix Figure A23.

To examine the selection of teachers into the transfer market, we first look at four cohorts, 2010-2013, such that we can follow them for five years. We further restrict attention to those for whom we can measure productivity, leaving us with 553 teachers who entered the state's data during those years. Of those, 207 applied to transfer at some point during the first five years. Only 124 remain in their original school and have not applied to transfer within five years of entering the district. The remaining 287 leave the district. Appendix Table A19 shows that there is very little difference in comparative advantage between teachers who applied to transfer and the teachers who did not. Teachers who apply for transfer have lower and less variable absolute advantage.

Accordingly, it is unlikely that the difference in per-student potential gains is due to teacher selection into transferring, particularly with regard to comparative disadvantage (with disadvantaged students). It is possible that the small differences in absolute disadvantage interacted with the under-representation of large classes accounts for some of the gap. The clearest selection into the transferring market, however, comes from the over-representation of schools with a high concentration of disadvantaged students. With a limited distribution of schools, there is less room to realize the gains from teachers' comparative advantages.

F Principal preferences estimation

We estimate principal preferences via maximum simulated likelihood, where we simulate from the normal distributions of the random effect at the level of the position-year. Let n index each simulation iteration and let $B_{jptn}(\theta)$ be the model-predicted probability that p rates j positively in year t in simulation iteration n at parameter vector θ. For each position p in year t, we construct the simulated likelihood as:

$$L_{pt} = \frac{1}{100} \sum_{n=1}^{100} \prod_{j \in \mathcal{J}_{pt}} (b_{jpt} B_{jptn}(\theta) + (1 - b_{jpt})(1 - B_{jptn}(\theta))), \tag{A19}$$

where $\mathcal{J}_{pt}$ is the set of teachers who applied to a position p in year t and b_{jpt} is an indicator for whether p rated j positively in the data. Our full simulated log likelihood function is:

$$l = \frac{1}{P}\sum_{p} logL_{pt}. \tag{A20}$$

Table A1: Applications: sample and summary statistics

	Full Sample	Elementary Sample	Value-Added Sample	2015 Sample
Applications				
N	2,163,711	337,754	13,819	2,702
On-Cycle	0.68	1.00	1.00	1.00
Instructional	0.70	1.00	1.00	1.00
Elementary	0.39	1.00	1.00	1.00
Applicants				
N	104,795	14,864	867	178
Female		0.92	0.87	0.89
Black		0.24	0.30	0.25
Hispanic		0.03	0.01	0.03
In-District		0.12	0.43	0.44
Choice Set Size		159.10	151.14	151.35
Application Rate		0.18	0.11	0.10
Transferred		0.23	0.43	0.51
Mean Commute Time		17.78	22.57	22.50
Experience		5.81	9.22	9.89
VA Econ Adv		-0.03	-0.03	-0.04
VA Econ Disadv		-0.02	-0.02	-0.03
Abs Adv		-0.03	-0.03	-0.03
Comp Adv in Econ Disadv		0.01	0.01	0.01
Positions				
N	38,921	1,824	1,784	296
Choice Set Size		1,293.54	71.89	88.63
Application Rate		0.14	0.11	0.10
Mean Class Size		26.40	26.40	25.69
Frac Econ Disadv		0.65	0.65	0.68
Frac Black		0.43	0.43	0.45
Frac Hispanic		0.24	0.24	0.25

The table shows count or mean statistics across different samples. The "Full Sample" includes all of the raw data, the "Elementary Sample" restricts to on-cycle elementary school instructional positions without specialization, the "Value-Added Sample" further restricts to teachers with value-added forecasts based on prior years, and the "2015 Sample" further restricts to the 2015 application cycle (for positions in the 2016 school year). We use the "Elementary Sample" for estimating principal preferences, the "Value-Added Sample" for estimating teacher preferences, and the "2015 Sample" for estimating counterfactual allocations. We do not include mean statistics for applicants and positions for the complete sample because we built the data on the subsample. Commute time is measured in minutes, absolute advantage is value-added at the representative school in the district, and choice set size is the number of positions in a teacher's choice set (Applicants panel) or the number of teachers with the position in their choice set (Positions panel).

Table A2: Relationship between Teacher Characteristics and Teacher Value-Added

	VA Mean	VA Adv	VA Disadv
Experience 1-2	0.0797	0.0744	0.0816
	(0.0326)	(0.0315)	(0.0334)
Experience 3-5	0.134	0.123	0.138
	(0.0322)	(0.0312)	(0.0331)
Experience 6-12	0.139	0.126	0.144
	(0.0320)	(0.0310)	(0.0329)
Experience 13-20	0.137	0.125	0.142
	(0.0320)	(0.0310)	(0.0329)
Experience 21-27	0.149	0.138	0.155
	(0.0322)	(0.0312)	(0.0331)
Experience 28+	0.132	0.121	0.135
	(0.0324)	(0.0314)	(0.0333)
Graduate degree	0.00263	0.00442	0.000950
	(0.00364)	(0.00352)	(0.00373)
Regular license	0.0531	0.0443	0.0574
	(0.0183)	(0.0177)	(0.0188)
NBPTS certified	0.0303	0.0303	0.0307
	(0.00528)	(0.00511)	(0.00542)
Praxis	0.00414	0.00573	0.00323
	(0.00241)	(0.00233)	(0.00247)
Mean DV	-0.00366	-0.0130	0.000960
R squared	0.0228	0.0219	0.0232
N	7335	7335	7335

The table shows the relationship between teacher characteristics and value added across student types (“Adv” and “Disadv”) or mean value added. The omitted experience category is having no experience

Table A3: Students and teachers: summary statistics for 2015-16

	Focal, 0-70%	Focal, 70-100%	Other, 0-70%	Other, 70-100%
Students				
Male (%)	50.76	51.03	51.18	51.46
White (%)	52.08	6.10	63.39	30.11
Black (%)	25.12	53.44	15.08	37.74
Hispanic (%)	12.58	34.00	12.86	23.72
Other minorities (%)	10.22	6.46	8.66	8.43
Student performance (level scores)				
Math	0.49	-0.18	0.17	-0.31
Student performance (gain scores)				
Math	0.06	0.07	-0.01	0.01
Teachers				
Experience (% of teachers)				
0 years	4.50	12.30	3.71	5.89
1-2 years	9.89	18.63	7.73	11.53
3-5 years	16.73	20.43	11.75	14.02
6-12 years	30.58	20.43	27.61	26.79
13-20 years	21.58	17.54	26.91	21.09
21-27 years	9.71	3.25	11.99	9.56
28 or more years	7.01	7.41	10.30	11.12
Graduate degree (%)	47.03	41.82	38.44	37.79
Regular license (%)	96.40	88.97	97.96	93.07
NBPTS certified (%)	15.11	5.42	13.61	7.59
Praxis score	0.32	0.02	0.28	0.05
Attrition rate (%)				
From school	14.03	27.31	16.50	23.69
From district	8.09	15.55	10.55	15.75
Schools				
Economically disadvantaged (%)	31.23	99.50	45.05	94.23
Mean math value-added				
Baseline	0.01	0.03	-0.01	0.00
Homogeneous	0.01	0.01	-0.01	-0.02
Using school means	0.17	0.17	0.11	0.13
Using alternative FEs	0.04	0.05	0.02	0.02

The table shows mean characteristics for students and teachers in our sample for the 2015-16 school year. Schools are split into whether they are in our focal district ("Focal") or in the rest of North Carolina ("Other") and whether more than 70% of the students in the school are economically disadvantaged. Math scores are standardized to have mean 0 and standard deviation 1 at the state-grade-year level. The alternate VA estimators are a (a) homogeneous value-added model with constant effects across student types, (b) a model that uses school mean characteristics rather than school fixed effects, and (c) a model that uses teacher-year fixed effects, rather than teacher-class-student type fixed effects, in the first residualization step.

Table A4: Teacher Value-Added Structural Parameters

	Estimates	Standard Errors	95% CI Lower Bound	95% CI Upper Bound
$\sigma_{\varepsilon 0}$	0.450	0.000	0.456	0.457
$\sigma_{\varepsilon 1}$	0.470	0.000	0.477	0.479
$\sigma_{\theta 0}$	0.110	0.007	0.108	0.137
$\sigma_{\theta 1}$	0.088	0.015	0.089	0.143
correlation($\theta_{c0t}, \theta_{c1t}$)	0.657	0.162	0.126	0.844
$\sigma_{\mu 0}$	0.249	0.007	0.262	0.284
$\sigma_{\mu 1}$	0.243	0.015	0.254	0.316
correlation(μ_{j0t}, μ_{j1t})	0.859	0.035	0.729	0.872

The table shows the estimates of a subset of the structural parameters of the production model – specifically the parameters corresponding to contemporaneous output. Non-disadvantaged students have index 1 while disadvantaged students have index 2. ε is the student-year idiosyncratic component, θ captures classroom effects, and μ describes a teacher's value-added. The remaining structural parameters describe the drift process of teacher value-added over time. Standard errors and confidence intervals are estimated with 100 bootstrap iterations.

Table A5: Potential Gains from Reassignment – Test Score Percentiles

	Per-Student Gains (σ)	As a Fraction of (Best-Actual)	Non-Disadvantaged	Disadvantaged
Alternate Allocations				
Best	0.050		0.089	0.016
Random	-0.003	-0.19	0.017	-0.021
Worst	-0.053	-3.63	-0.052	-0.054
Alternate Policies				
Best w/i School	0.012	0.98	0.015	0.010
Replace Bottom 5% of Teachers	0.011	0.74	0.012	0.010
Targeting Student Types				
Max Non-Disadvantaged VA	0.024	1.66	0.130	-0.072
Max Disadvantaged VA	0.016	1.12	-0.047	0.074

The table shows the potential gains from reassignments of teachers to different schools. Test scores are constructed as the raw score percentile (from 0 to 1), where percentiles are calculated for each grade-year in the state. We then normalize the test scores to be in standard deviation units based on the standard deviation of the uniform distribution. The sample is all teachers with non-missing value-added measures in 2016, along with their corresponding 2016 assignments. Gains come from better matching of teachers to students, as teachers' effectiveness may differ across student types. The first column shows the per-student gains from various allocations relative to the actual allocation. The second column shows the gain as a fraction of the full difference between the best (output-maximizing) and actual allocations. The third and fourth columns show the per-student gains, relative to the actual allocation, for non-disadvantaged and disadvantaged students. The best within school allocation only changes the teacher-classroom assignments within a school. "Replacing Bottom 5% of Teachers" refers to replacing the bottom 5% of teachers according to realized per-student output with teachers with median value-added for each student type. The targeting student types allocations are the ones that maximize per-student output for students of one type only. We assign classrooms the mean student composition and class sizes in that school in 2016 in all allocations except the "Best w/i School" and "Constant Class Size" allocations.

Table A6: Potential Gains from Reassignment – Constant Class Size

	Per-Student Gains (σ)	As a Fraction of (Best-Actual)	non-Disadvantaged	Disadvantaged
Alternate Allocations				
Best	0.021		0.018	0.023
Random	0.000	0.02	0.002	-0.001
Worst	-0.020	-0.94	-0.015	-0.023
Alternate Policies				
Best w/i School	0.003	0.16	0.007	0.000
Replace Bottom 5% of Teachers	0.017	0.82	0.023	0.012
Targeting Student Types				
Max Non-Disadvantaged VA	0.003	0.17	0.123	-0.090
Max Disadvantaged VA	0.005	0.22	-0.111	0.096

The table shows the potential gains from reassignments of teachers to different schools where each school has the same number (but possibly different composition) of students per class. The sample is all teachers with non-missing value-added measures in 2016, along with their corresponding 2016 assignments. Gains come from better matching of teachers to students, as teachers' effectiveness may differ across student types. The first column shows the per-student gains from various allocations relative to the actual allocation. Gains are measured in student standard deviations (σ). The second column shows the gain as a fraction of the full difference between the best (output-maximizing) and actual allocations. The third and fourth columns show the per-student gains, relative to the actual allocation, for non-disadvantaged and disadvantaged students. The best within school allocation only changes the teacher-classroom assignments within a school. "Replacing Bottom 5% of Teachers" refers to replacing the bottom 5% of teachers according to realized per-student output with teachers with median value-added for each student type. The targeting student types allocations are the ones that maximize per-student output for students of one type only. We assign classrooms the mean student composition and class sizes in that school in 2016 in all allocations except the "Best w/i School" and "Constant Class Size" allocations.

Table A7: Same-Race and Same-Gender Effects on Test Scores

	Student Res
Black Teacher - Black Student	0.00225
	(0.00164)
Hispanic Teacher - Hispanic Student	-0.00556
	(0.00549)
Female Teacher - Female Student	0.00478
	(0.000550)
Fixed Effects	Teacher, School
Mean DV	0.0000115
Clusters	37940
N	5158740

An observation is a student-year and the outcome is the student's math score residualized by student demographics including lagged scores, school fixed effects, and teacher experience measures. The regressors include measures of demographic match between student and teacher. The regression includes school fixed effects and teacher fixed effects. Standard errors are clustered at the teacher level.

Table A8: Teacher Value-Added Structural Parameters with Alternate Forms of Heterogeneity

	Race	Achievement
$\sigma_{\varepsilon 0}$	0.465	0.481
$\sigma_{\varepsilon 1}$	0.457	0.439
$\sigma_{\theta 0}$	0.091	0.099
$\sigma_{\theta 1}$	0.110	0.102
correlation($\theta_{c0t}, \theta_{c1t}$)	0.637	0.628
$\sigma_{\mu 0}$	0.233	0.240
$\sigma_{\mu 1}$	0.261	0.282
correlation(μ_{j0t}, μ_{j1t})	0.900	0.844

The table shows the estimates of a subset of the structural parameters of production models with alternate forms of heterogeneous teacher effects – specifically by race and prior achievement. In the first column, non-white students have index 1 while White students have index 2. In the second column, students with below median prior math achievement have index 1 while students with above median prior math achievement have index 2. ε is the student-year idiosyncratic component, θ captures classroom effects, and μ describes a teacher's value-added. The remaining structural parameters describe the drift process of teacher value-added over time.

Table A9: Application timing

	Obs	Mean days	Median days	Share 0 days
Stock	196,779	3.6	0	0.72
Flow	146,382	2.1	0	0.75

(a) Wait times until applying

	Obs	Mean fraction of days	Mean fraction of applications	Mean days since posting
First day	14,864	0.61	0.65	23.47
Subsequent days	40,850	0.14	0.13	11.55

(b) First day versus subsequent days

	Obs	April or before	May	June	July	August
First day (all teachers)	14,864	0.20	0.25	0.22	0.18	0.15
Last day (all teachers)	14,864	0.09	0.15	0.21	0.26	0.29
First day (transfers)	2,547	0.27	0.30	0.24	0.14	0.05
Last day (transfers)	2,547	0.10	0.17	0.25	0.29	0.19

(c) Timing of first and last days

The tables show statistics related to application timing. Panel (a) shows how long it took an applicant to apply to positions that were in "stock" (already posted) on the day the teacher first applied on the platform or in "flow" (posted after the day the teacher first applied on the platform). Panel (b) shows application statistics for the first day a teacher applied on the platform in a cycle versus subsequent days. "Mean days since posting" is the mean number of days a vacancy had been posted at the time the teacher applied. Panel (c) shows the (monthly) timing of when an applicant's first and last application days of the cycle occurred. "All teachers" includes all applicants while "transfers" includes just teachers who ended up in new schools.

Table A10: Pseudo R-squareds for principal rating models

	Non-Title I	Title I
Demographics	0.008	0.004
Teacher Characteristics	0.031	0.018
Value Added	0.006	0.003
EVAAS	0.000	0.000
Demographics + Teacher Characteristics	0.039	0.023
Demographics + Value Added	0.016	0.006
Teacher Characteristics + Value Added	0.033	0.020
EVAAS + Value Added	0.007	0.003
Demographics + Teacher Characteristics + Value Added	0.041	0.025
Demographics + Teacher Characteristics + EVAAS + Value Added	0.041	0.025

The table shows pseudo R-squareds from logit models for whether a principal rates an application highly (a positive rating, an interview, or an offer). Each model includes position fixed effects. The pseudo R-squared is the percentage improvement in the likelihood relative to a model with only the fixed effects. Demographics are measures of the teacher's race and gender, interacted with the school's racial composition. Teacher characteristics are experience, licensing, certification, and Praxis scores. Value Added is our model's forecast of the teacher's causal effect on student test scores from the assignment. EVAAS is the measure of teacher performance that the state uses and released to teachers.

Table A11: Multi-classroom teacher prevalence

Year	All	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
2012	0.264	0.109	0.187	0.618	0.621	0.631
2013	0.287	0.124	0.210	0.636	0.631	0.649
2014	0.300	0.152	0.227	0.633	0.625	0.644
2015	0.363	0.256	0.345	0.615	0.598	0.602
2016	0.391	0.305	0.392	0.595	0.591	0.595
2017	0.385	0.291	0.399	0.612	0.569	0.596
2018	0.393	0.307	0.425	0.596	0.586	0.578
Estimation sample	0.417					

The table shows the prevalence of teachers having multiple classrooms, separately by teacher's grade and year. The sample includes teachers for whom we can calculate math value-added. Our estimation sample consists of teachers, with value-added forecasts, who applied to elementary school positions.

Table A12: Estimated Experience Returns to Teacher Value-Added

	1	2	3	4	5	6	7+
Estimate	0.056	0.077	0.083	0.088	0.088	0.091	0.070
Standard Error	0.004	0.004	0.005	0.005	0.005	0.005	0.005

The table shows the estimated experience returns for math test scores, where the scores have been normalized to have mean 0 and standard deviation 1 for students in a given grade-year. Columns designate the number of prior years of experience. The omitted category is teachers with no prior experience.

Table A13: Predicting Student Residuals by Teacher Type

	Student res	Student res
Share disadvantaged	-0.0549	-0.0409
	(0.0251)	(0.0202)
Share disadvantaged x CA in disadvantaged	0.0820	0.0697
	(0.0356)	(0.0283)
Num teachers	3214	3214
Num students	157671	157671
Mean CA	-0.00805	-0.00805
SD CA	0.0624	0.0624
Controls	No	Yes

The table assesses whether changes in the share of economically disadvantaged students predict changes in student test score residuals differently by teacher type across transfers. Teacher type is defined by comparative advantage in pre-transfer schools, with "CA in disadvantaged" an indicator for whether the teacher is above median in comparative advantage in teaching disadvantaged students. The outcome is changes in average teacher-by-school student residuals across transfers. "Share disadvantaged" is the change in the average share of economically disadvantaged students teacher j taught when moving from one school to another. Controls include a cubic in average experience in the school, an indicator for experience missingness, and transfer year indicators. Standard errors are clustered at the teacher level.

Table A14: Autocorrelations in class size

Class size, school level

Variables	Class Size t	Class Size t-1	Class Size t-2	Class Size t-3	Class Size t-4
Class Size t	1.0000				
Class Size t-1	0.7329	1.0000			
	(0.0000)				
Class Size t-2	0.6248	0.6966	1.0000		
	(0.0000)	(0.0000)			
Class Size t-3	0.4093	0.5261	0.6598	1.0000	
	(0.0000)	(0.0000)	(0.0000)		
Class Size t-4	0.3722	0.3746	0.4365	0.5796	1.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	

Nb. obs. : 247

Class size, teacher level

Variables	(Res.) Size t	(Res.) Size t-1	(Res.) Size t-2	(Res.) Size t-3	(Res.) Size t-4
(Res.) Size t	1.0000				
(Res.) Size t-1	0.3668	1.0000			
	(0.0000)				
(Res.) Size t-2	0.2688	0.3717	1.0000		
	(0.0000)	(0.0000)			
(Res.) Size t-3	0.2900	0.1272	0.2699	1.0000	
	(0.0000)	(0.0186)	(0.0000)		
(Res.) Size t-4	0.1173	0.1438	0.0698	0.3098	1.0000
	(0.0301)	(0.0077)	(0.1978)	(0.0000)	

Nb. obs. : 342

The table shows correlations (within unit) between class size in one year and class size in a prior year. In the top panel, a unit of analysis is a school and class size is the mean across all of the school's classrooms (that generate math test scores). In the bottom panel, a unit of analysis is a teacher and class size is residualized by school-year fixed effects such that residual class size compares how a teacher's class size deviates from the school-year mean.

Table A15: Autocorrelations in class composition

Class composition, school level

Variables	Frac Disadv t	Frac Disadv t-1	Frac Disadv t-2	Frac Disadv t-3	Frac Disadv t-4
Frac Disadv t	1.0000				
Frac Disadv t-1	0.9602	1.0000			
	(0.0000)				
Frac Disadv t-2	0.9430	0.9555	1.0000		
	(0.0000)	(0.0000)			
Frac Disadv t-3	0.9363	0.9370	0.9496	1.0000	
	(0.0000)	(0.0000)	(0.0000)		
Frac Disadv t-4	0.9435	0.9467	0.9554	0.9775	1.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	

Nb. obs. : 247

Class composition, teacher level

Variables	(Res.) Dis t	(Res.) Dis t-1	(Res.) Dis t-2	(Res.) Dis t-3	(Res.) Dis t-4
(Res.) Dis t	1.0000				
(Res.) Dis t-1	0.3170	1.0000			
	(0.0000)				
(Res.) Dis t-2	0.2898	0.3200	1.0000		
	(0.0000)	(0.0000)			
(Res.) Dis t-3	0.1524	0.2076	0.3723	1.0000	
	(0.0047)	(0.0001)	(0.0000)		
(Res.) Dis t-4	0.0921	0.0512	0.2203	0.3925	1.0000
	(0.0889)	(0.3450)	(0.0000)	(0.0000)	

Nb. obs. : 342

The table shows correlations (within unit) between class composition (fraction of students that are economically disadvantaged) in one year and class composition in a prior year. In the top panel, a unit of analysis is a school and class composition is the (weighted) mean across all of the school's classrooms (that generate math test scores). In the bottom panel, a unit of analysis is a teacher and class composition is residualized by school-year fixed effects such that residual class composition compares how a teacher's class composition deviates from the school-year mean.

Table A16: Teacher experience and student assignment

	(1) Outcome	(2) Outcome	(3) Outcome	(4) Outcome	(5) Outcome
Outcome: Share economically disadvantaged students assigned					
log(experience)	-0.0369	-0.0311	-0.0063	-0.0029	-0.0021
	(0.0013)	(0.0028)	(0.0005)	(0.0011)	(0.0011)
Outcome: Share Black students assigned					
log(experience)	-0.0331	-0.0195	-0.0010	-0.0008	-0.0005
	(0.0010)	(0.0023)	(0.0004)	(0.0008)	(0.0010)
Outcome: Average student lagged math score					
log(experience)	0.0887	0.0474	0.0461	0.0173	0.0115
	(0.0023)	(0.0049)	(0.0016)	(0.0033)	(0.0041)
Outcome: Share gifted status					
log(experience)	0.0231	0.0106	0.0161	0.0053	0.0074
	(0.0007)	(0.0014)	(0.0006)	(0.0012)	(0.0016)
New only		X		X	X
Year FE			X	X	
School FE			X	X	
School-year FE					X
N	1,879,666	258,723	1,879,666	258,723	258,723

Standard errors in parentheses.

The table shows separate regression results for different outcomes on the log of a teacher's prior experience. Outcomes are mean characteristics of the students in a teacher's classroom. "New only" indicates that the sample only includes teachers new to the school; thus, the regression compares outcomes across teachers new to the school depending on the teacher's experience.

Table A17: Applications to Title I and non-Title schools

	Obs	Mean choice set	Median	Mean prob.	25th	50th	75th	Std. dev.	Overall mean prob.
Title I	14,747	85.3	68	0.176	0.010	0.056	0.264	0.237	0.137
non-Title I	14,747	74.0	66	0.176	0.013	0.084	0.270	0.217	0.155
Gap	14,747			-0.001	-0.049	0.003	0.041	0.134	-0.018

The table shows application statistics to positions at Title I and non-Title I schools. Columns (2) and (3) show the mean and median choice set sizes for an applicant. “Gap” shows the difference in statistics across the two school types.

Table A18: Predicting posted positions

	(1) Positions	(2) Positions
Class size	-0.0199	
	(0.0125)	
Fraction disadvantaged		1.503
		(0.544)
N	116	116

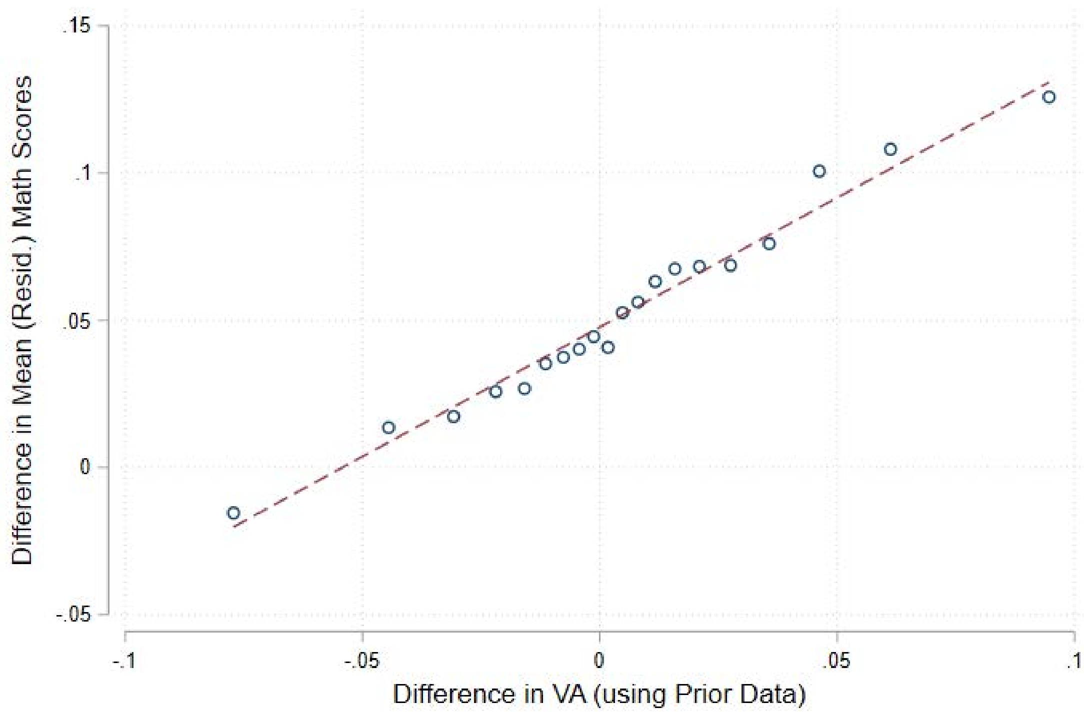
An observation is a school-year. The outcome is the number of positions posted in an application cycle and the regressors are characteristics of the school's mean class. Robust standard errors are in parentheses.

Table A19: Transferring and non-transferring teachers' value added

	(1) Did not apply			(2) Applied to transfer		
	mean	sd	count	mean	sd	count
Comparative advantage	0.0001	0.0351	528	-0.0002	0.0367	506
Absolute advantage	0.0034	0.1210	528	0.0219	0.1508	506

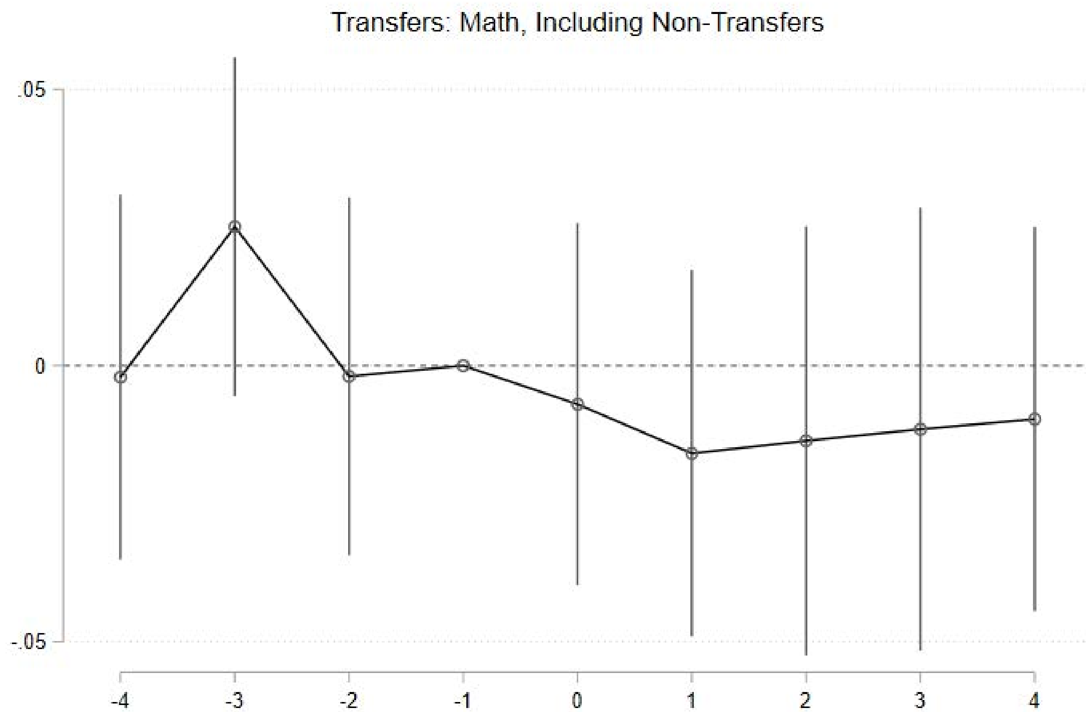
The table shows the means and standard deviations of absolute and comparative advantage for teaching economically advantaged students by whether the teacher ever submits an application to transfer. An observation is a teacher with a value-added forecast. These are pooled over years 2010 through 2018.

Figure A1: Math Comparative Advantage Forecast Unbiasedness



The figure is a binscatter, where an observation is a teacher-year and "Difference in VA" is the difference in a teacher's math value-added between economically disadvantaged and advantaged students. Value-added estimates are predictions using data from prior years. Units are student standard deviations. The y-axis is the difference in mean student math test score, residualized by student demographics including lagged scores, school fixed effects, and teacher experience measures. The mean is taken over all students (of a given type) for a given teacher-year and the difference is between a teacher's economically disadvantaged and advantaged students.

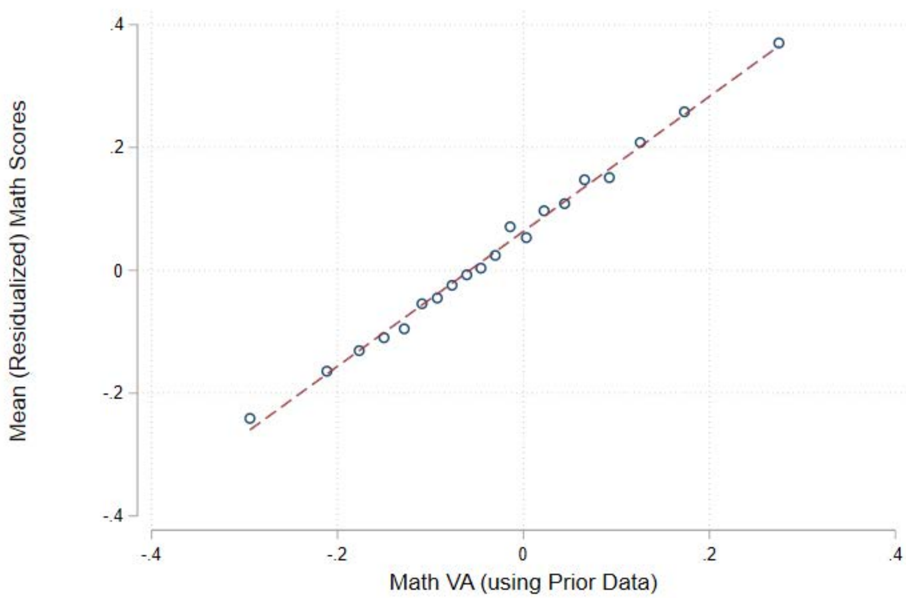
Figure A2: Transfer event study



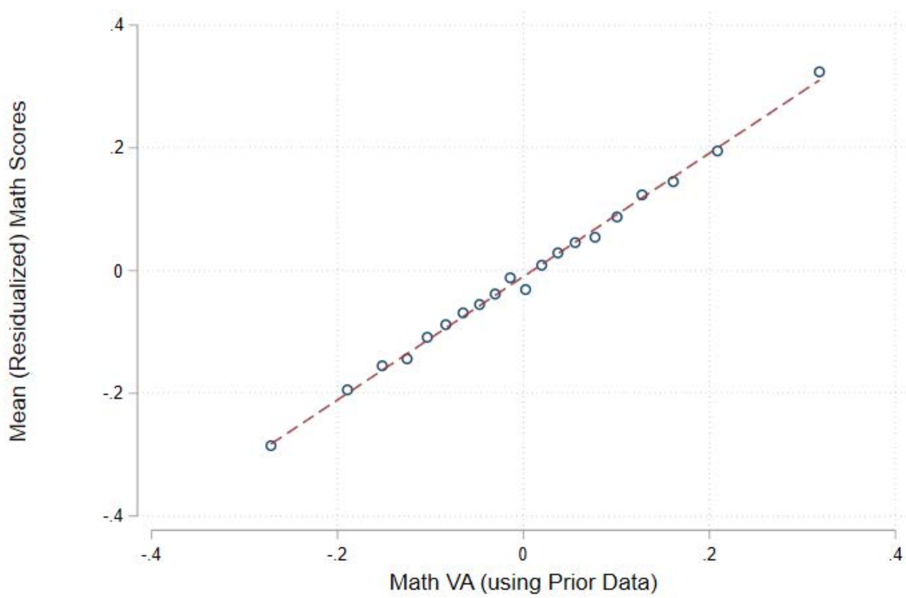
The figure shows event study coefficient estimates and 95% confidence intervals. The outcome is residualized math test score (residualized by student observables including lagged scores, school fixed effects, and an experience function), in student standard deviation units. The event is the teacher's first transfer from one school to another school in the state, where non-transfers do not have an event. We include teacher and year fixed effects and follow Sun and Abraham (2021) in constructing the estimates.

Figure A3: Forecast Unbiasedness for Large Changes in Class Size

(a) Large decreases in class size

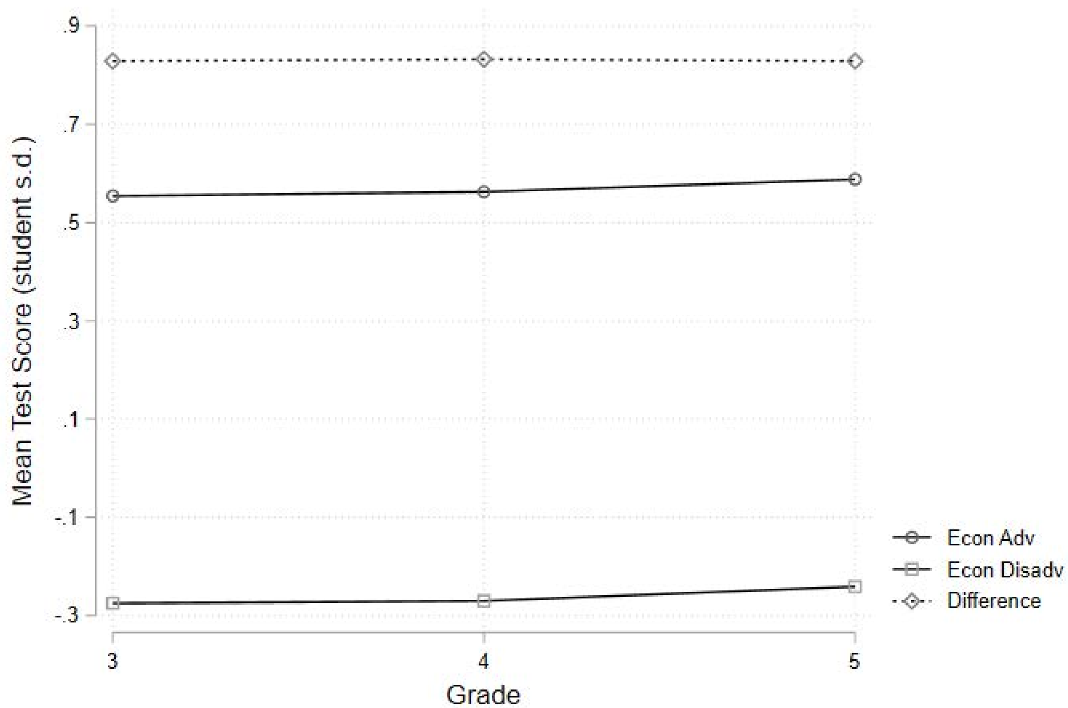


(b) Large increases in class size



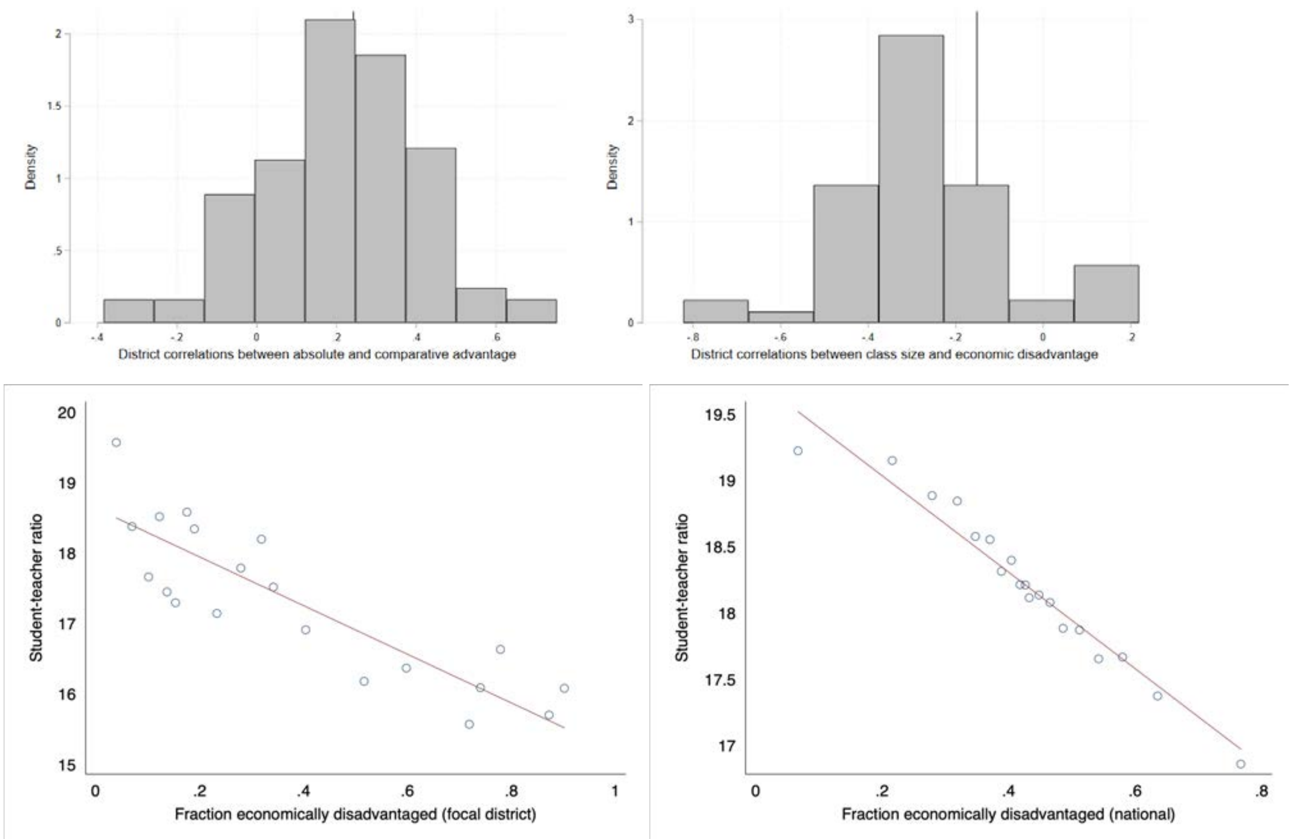
The figure shows a binscatter of student residual test scores by value-added prediction where an observation is a teacher-year. For decreases, the sample consists of all teachers where the class size used for prediction exceeds the class size in the target by more than 10 students. For increases, the sample consists of all teachers where the class size used for prediction is less than the class size in the target by more than 10 students.

Figure A4: Test Scores by Student Type, across Grades



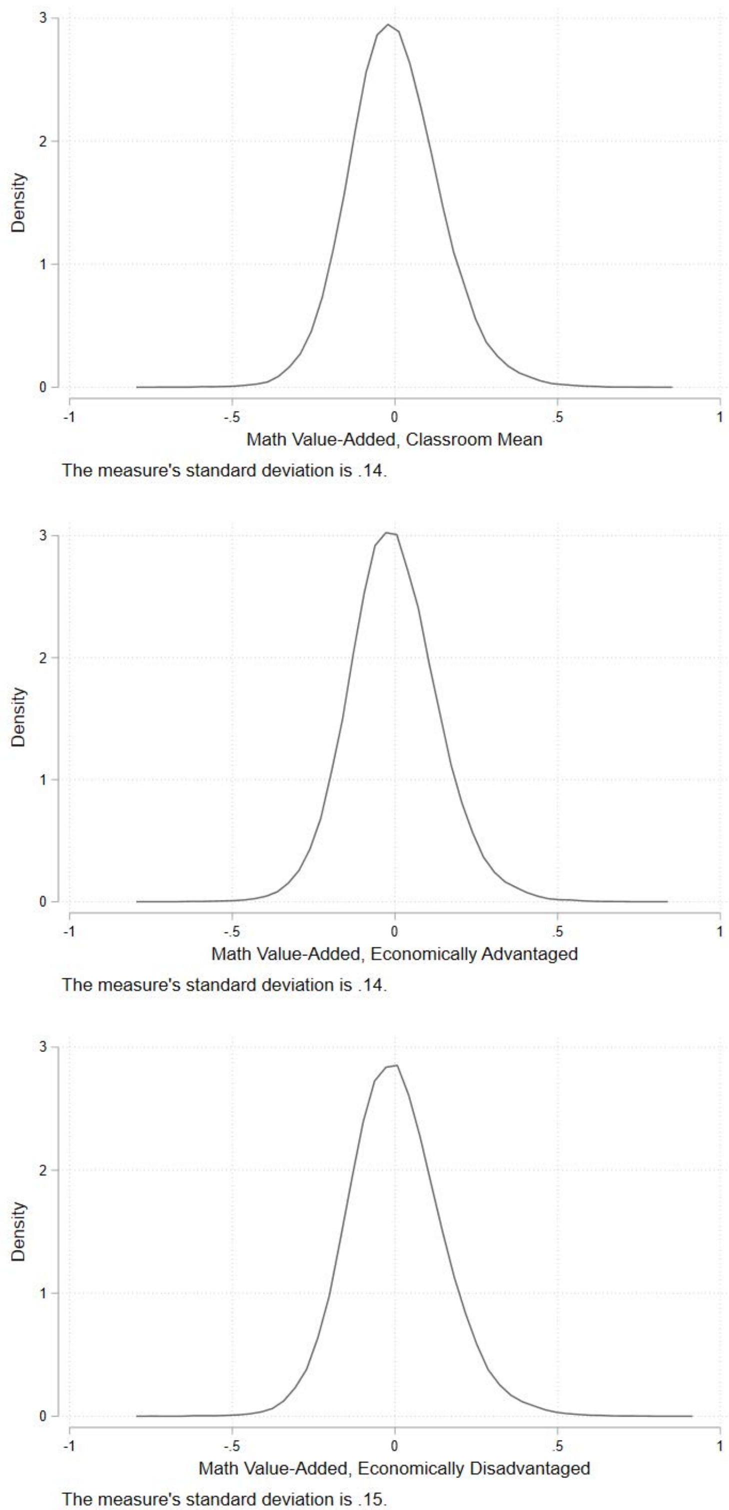
The figure plots mean student test scores in our focal district for economically advantaged and disadvantaged students and the difference between the groups. Test scores are in student standard deviation units and are plotted across elementary school grades.

Figure A5: External validity of assignment features



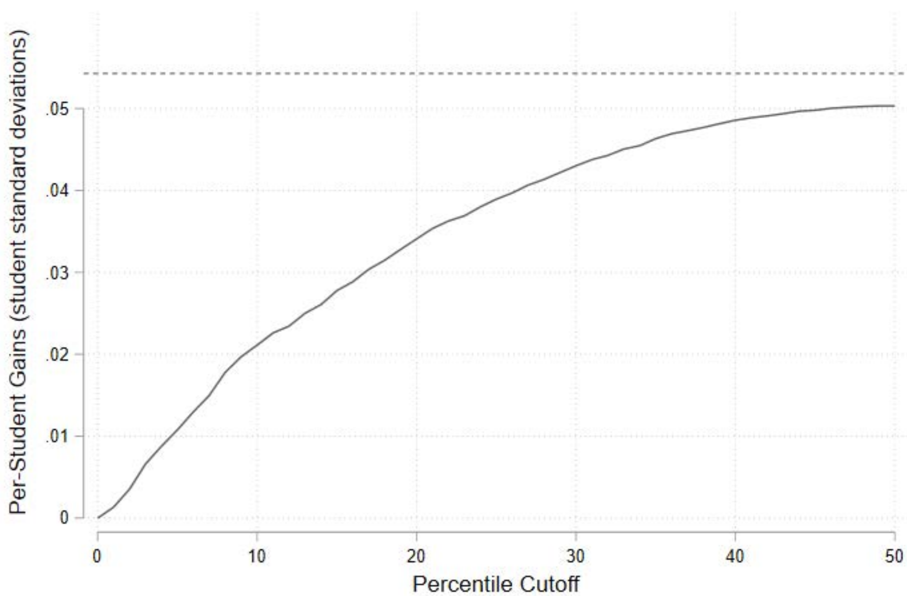
The figures show how the features of our assignment problem extend to other districts. In the top row, we calculate the correlation between absolute advantage and comparative advantage in teaching economically disadvantaged students (left) and the correlation between class size and the economically disadvantaged share (right). An observation is a district in North Carolina. The vertical line shows the correlation at our focal district. In the bottom row, we use national data from the Common Core of Data and restrict to elementary schools in 2012-13. The two figures are binscatters of the within-district relationship between a school's fraction economically disadvantaged and its student-teacher ratio. The figure on the left is for our focal district while the figure on the right includes all schools in the US, where an observation is a school and district fixed effects are included.

Figure A6: Value-Added Distribution

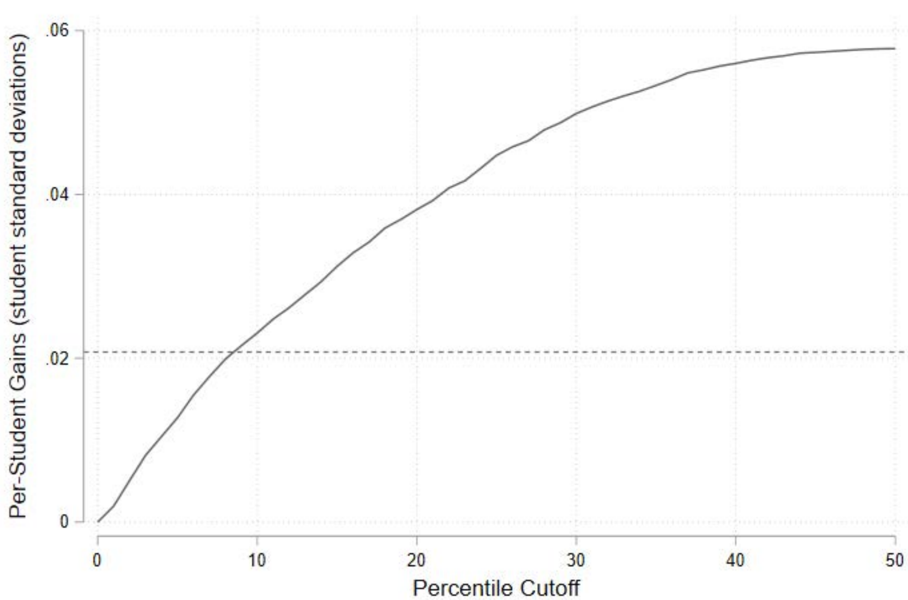


The figures show kernel density plots of our forecast of a teacher's value-added in a given year at the school they actually teach at (panel A), for economically advantaged students (panel B), and for economically disadvantaged students (panel C). The forecast uses only data from prior years. The units are student standard deviations.

Figure A7: Gains from teacher replacement



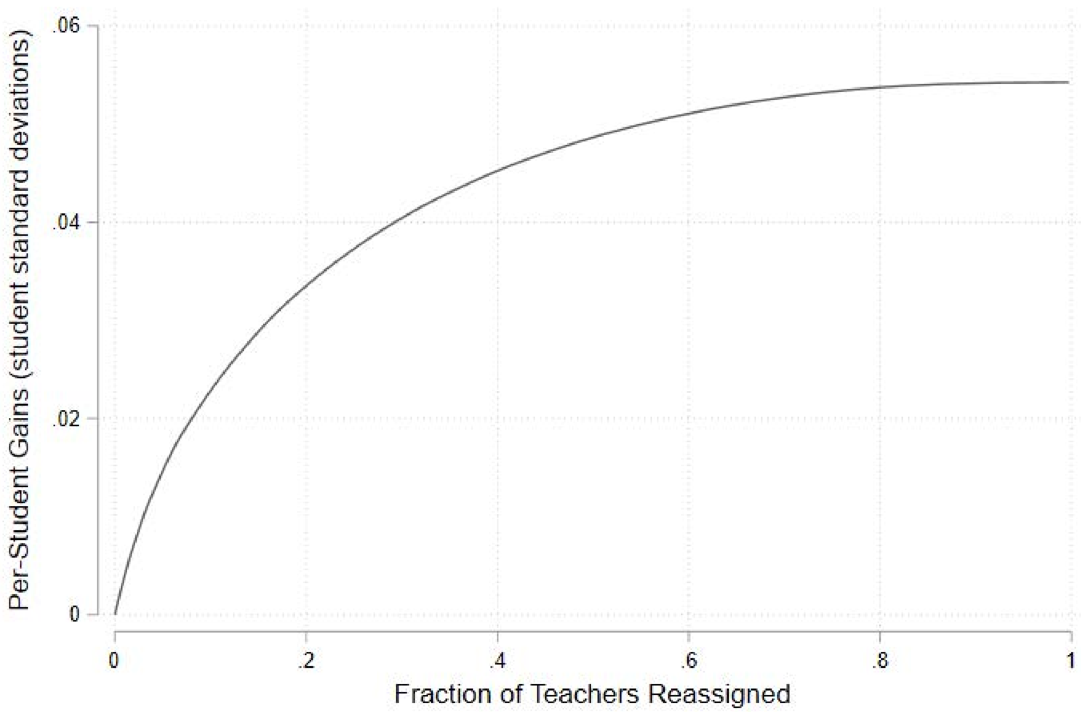
With class size variation



Constant class size

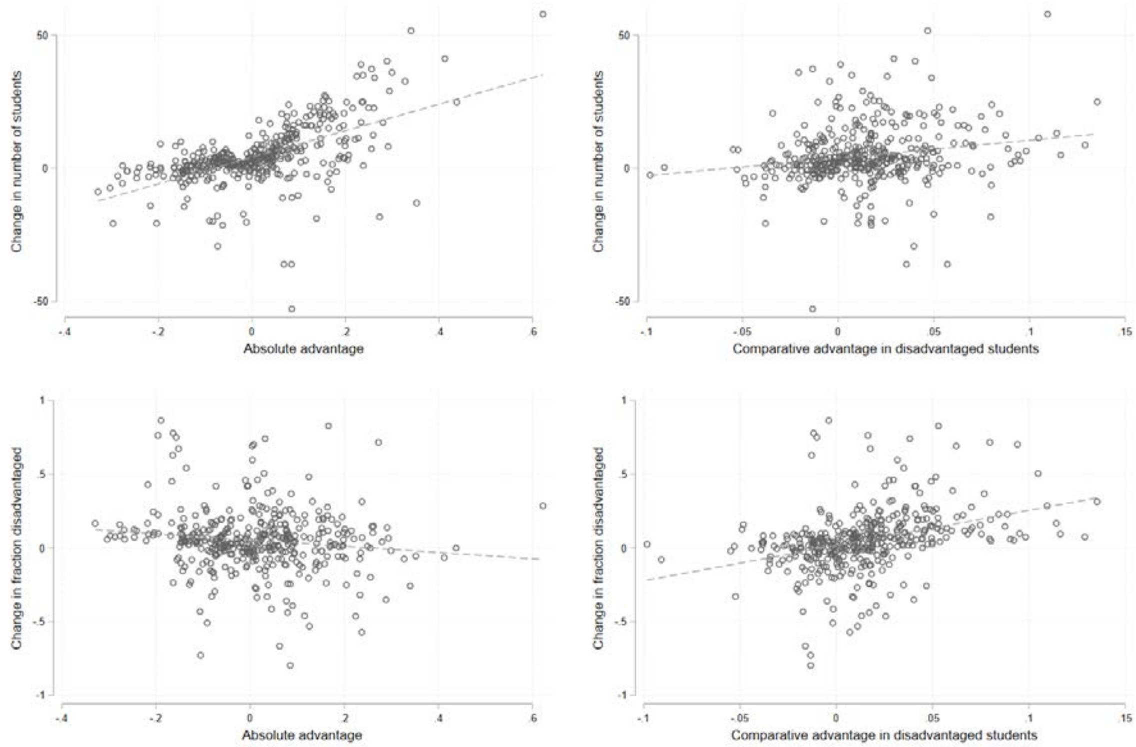
This Figure shows the results from policies that replace the X% of low-performing teachers with median value-added teachers, where the x-axis shows different values of X. The sample is the 2016 teachers with value-added forecasts. We assess performance based on realized value-added in the data (i.e., at the schools and classrooms a teacher is actually at in the data), and the median value-added teacher has median values for both dimensions of value-added. The y-axis is per-student gains in achievement. The top panel uses class size variation while the bottom panel imposes constant class sizes (at the district mean). The horizontal dashed lines are the gains from the output-maximizing allocation of existing teachers across schools in the district.

Figure A8: Student Gains by Fraction of Teachers Reassigned



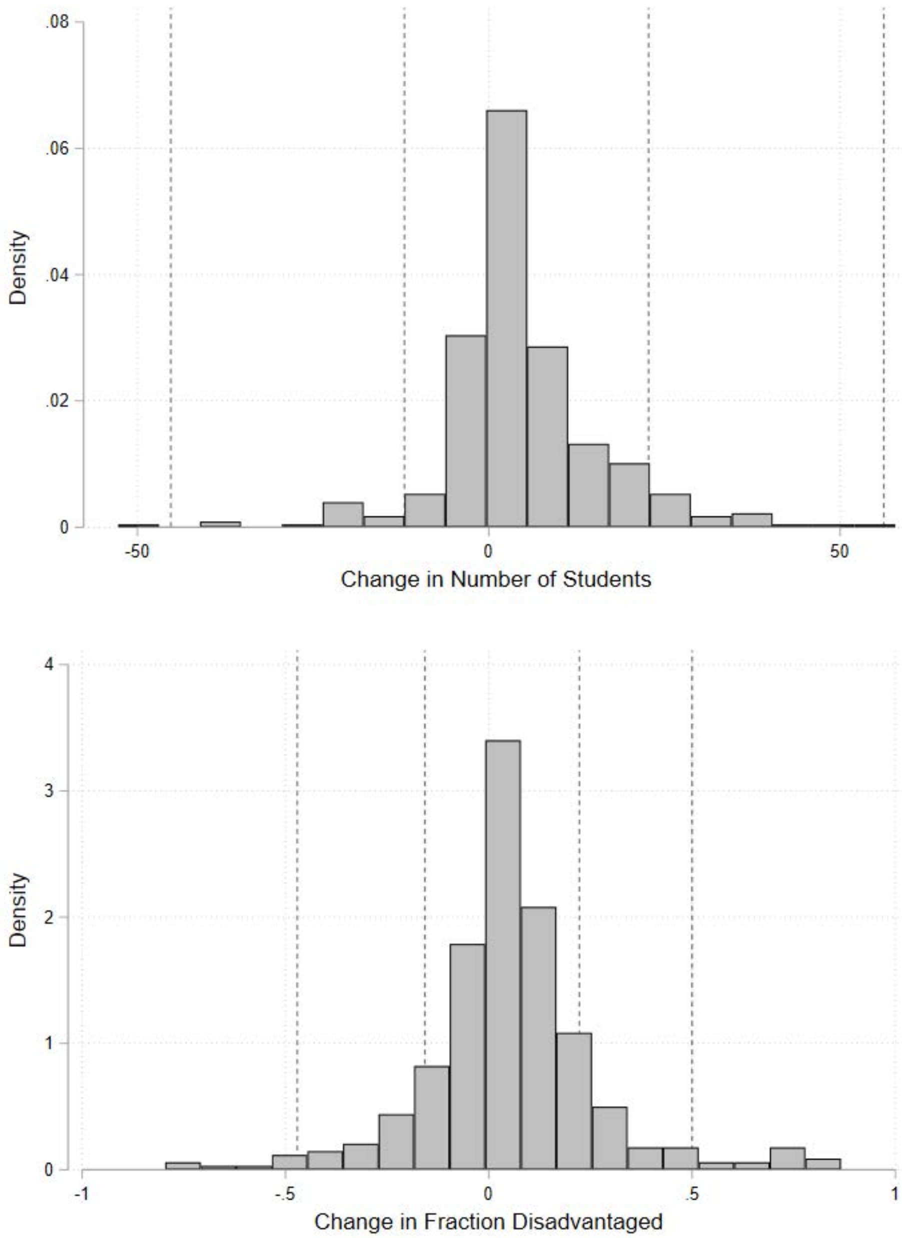
The figure plots the potential per-student math test score gains (in student standard deviation units) as a function of the fraction of teachers that are assigned to a school different than their actual school. The sample consists of the 2016 teachers with math value-added scores.

Figure A9: Changes in a teacher's classroom composition and size between the output-maximizing and actual allocations



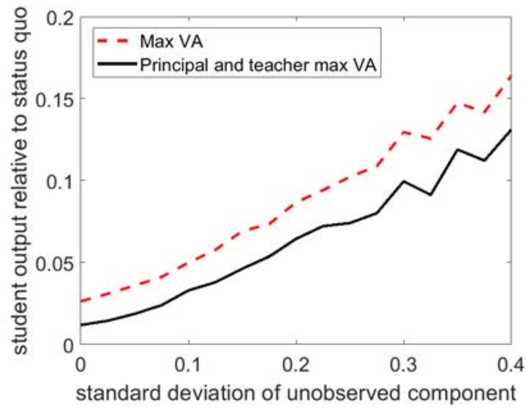
The figures show scatterplots and lines of best fit for the 2016 sample of teachers with value-added scores. In the top row, the variable of interest is the difference in the number of students a teacher teaches between the output-maximizing and actual allocations. Positive numbers are teachers who have more students in the output-maximizing allocation than in the actual. In the bottom row, the variable of interest is the difference in the fraction of disadvantaged students a teacher teaches between the output-maximizing and actual allocations. In the left column, teachers are ordered on the x-axis by absolute advantage (value-added at a representative school). In the right column, the teachers are sorted by comparative advantage in teaching economically disadvantaged students.

Figure A10: Optimal teacher placement relative to placement that generated value-added

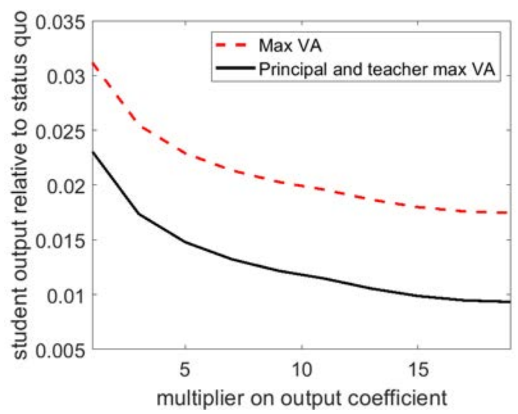


The figures show histograms for the 2016 sample of teachers with value-added scores. In the top panel, the variable of interest is the difference in the number of students a teacher teaches between the output-maximizing allocation and the classrooms that generated the teacher's value-added forecast. Positive numbers are teachers who have more students in the output-maximizing allocation than in the estimation data. In the bottom row, the variable of interest is the difference in the fraction of disadvantaged students a teacher teaches between the output-maximizing allocation and the classrooms that generated the teacher's value-added forecast. The vertical dashed lines represent the 1st, 10th, 90th, and 99th percentiles of the distribution we use for validation of our value-added measures in Table 1.

Figure A11: Simulations of two forms of misspecification



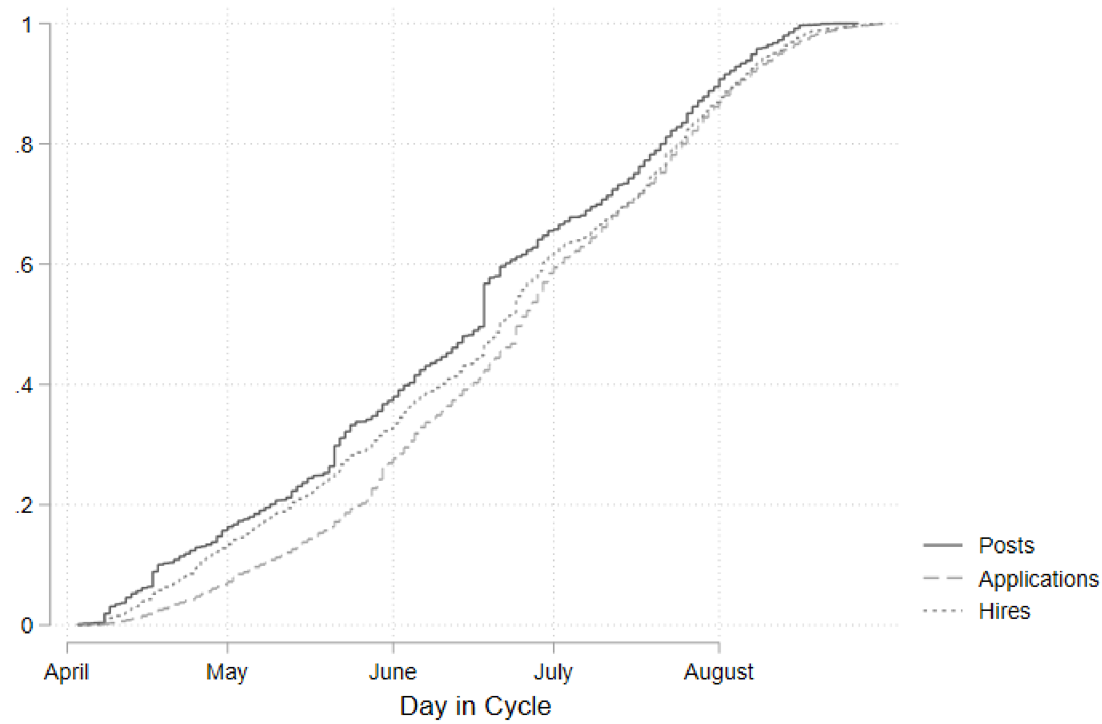
Unmodelled match effects



Attenuation in coefficient on output in teacher preferences

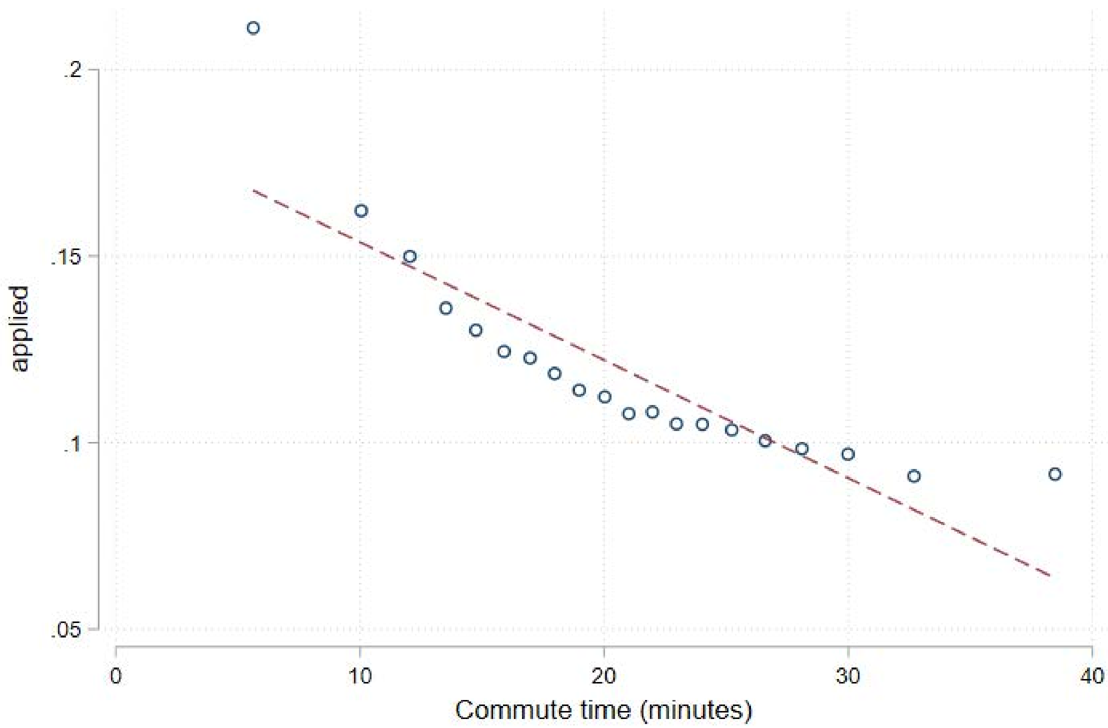
The figures show results from simulation exercises where we vary parameters related to match effects. In each figure the y-axis is the mean student achievement relative to the status quo. The dashed red line is the achievement in the output-maximizing allocation while the solid black line is the achievement in the equilibrium where principals and teachers each have preferences in order of value-added produced. The top panel adds an iid unobserved component to match effects, where the x-axis is the standard deviation of this component. The bottom panel varies the coefficient in teacher preferences on value-added. If our model misses match effects that teachers are aware of, then the preference coefficient might be attenuated. The x-axis in the bottom panel shows by how much we multiply our estimated coefficient on value-added.

Figure A12: Market timing



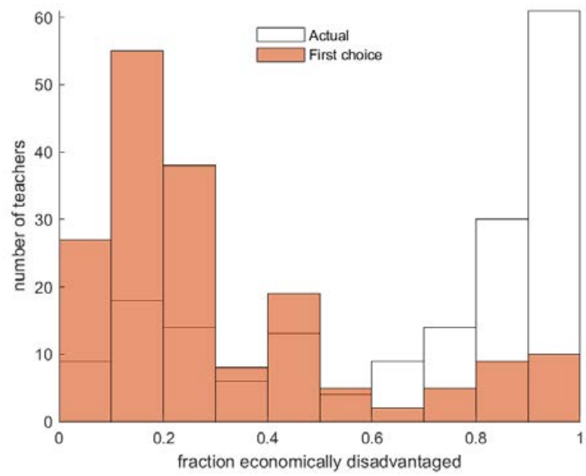
The figure shows CDFs for postings, applications, and hires (the application date of the application that led to a hire).

Figure A13: Application probabilities against commute time



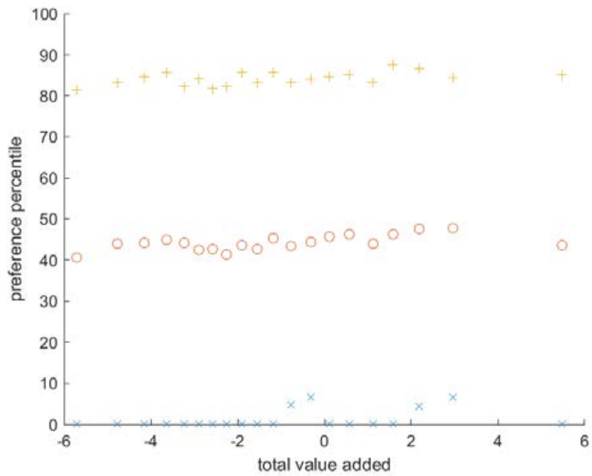
This Figure plots the probability of applying against commute time (measured in one-way minutes). The Figure residualizes for applicant fixed effects.

Figure A14: Number of teachers by fraction economically disadvantaged



This Figure plots histograms of the number of teachers, by the fraction of students who are economically disadvantaged. The histograms are for the actual positions in the data (in white) and the positions teachers would have if they could all have their top choice (in red).

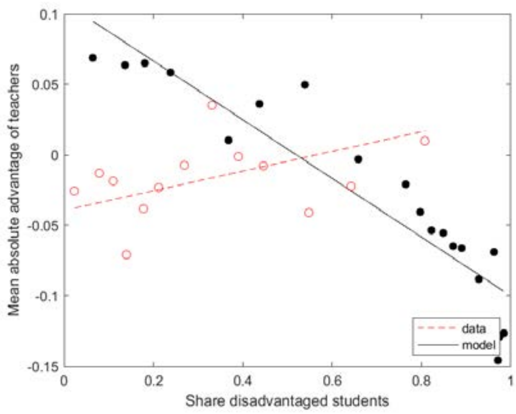
Figure A15: Bivariate preference relationship – principal model without output



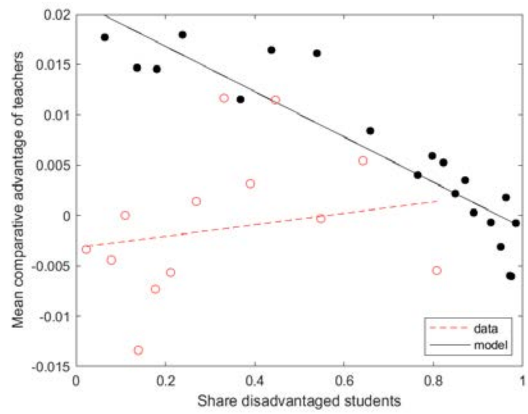
This figure shows a binscatter of the bivariate relationships between teacher output and principal preferences. We estimate each principal's ranking over teachers and order teachers from a principal's most preferred (100) to least preferred (0). The estimated model does not include value-added as a characteristic. The figure shows the bivariate relationship between the teacher's total value-added in the position and the mean preference percentile of the principal for the teacher in the principal preference model. The middle set of points (red circle) is the mean percentile, while the top (orange cross) and bottom (blue x) sets of points are the 10th and 90th percentiles, respectively.

Figure A16: Model fit: teacher serial dictatorship based on absolute advantage (descending)

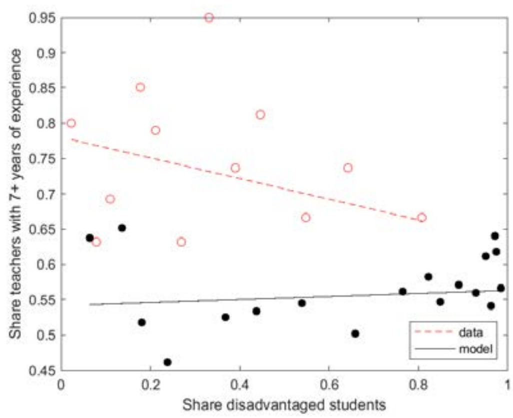
(a) Teacher absolute advantage



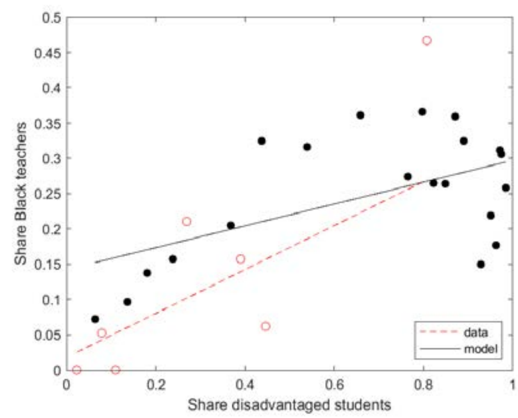
(b) Teacher comparative advantage



(c) Teachers with 7+ years of experience



(d) Teachers that are Black

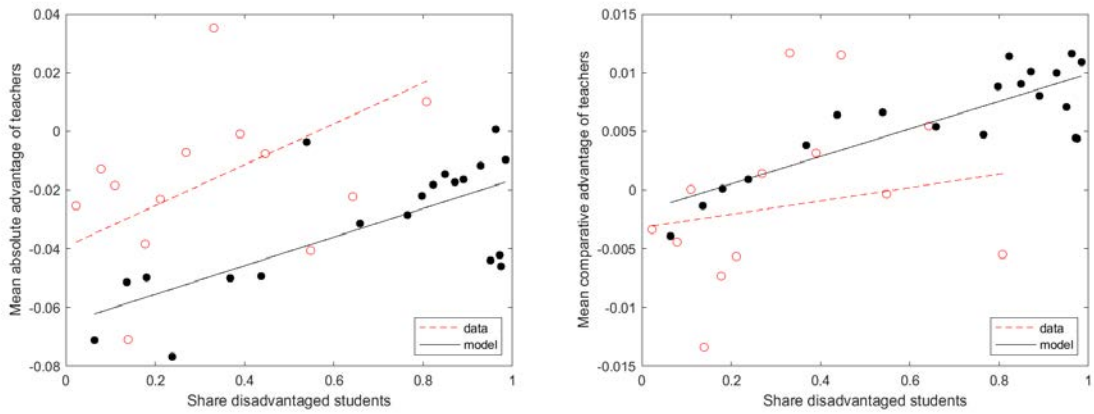


This Figure compares the allocations implied by a model in which the allocation is determined by a serial dictatorship where teachers go in descending order of their absolute advantage to the allocations we observe in the data. The data refers to all teachers in the district. The model refers to the teachers who apply in the transfer system for whom we have value-added scores.

Figure A17: Model fit: teacher serial dictatorship based on experience (descending)

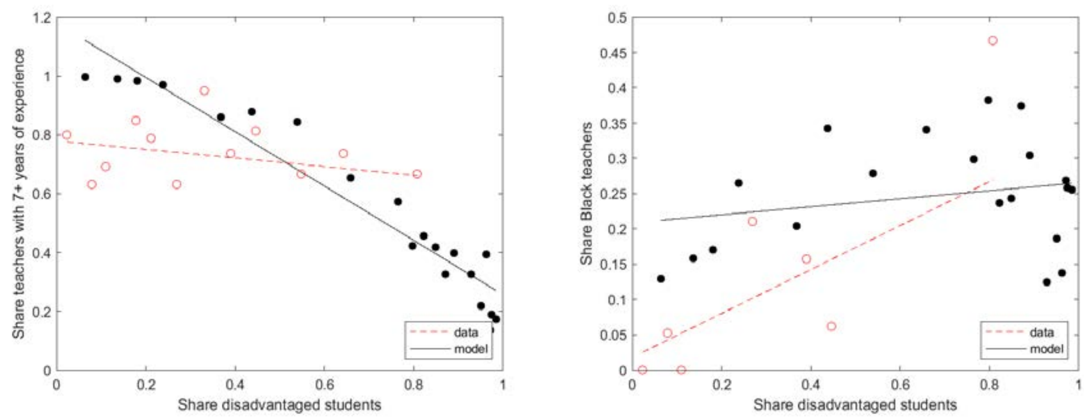
(a) Teacher absolute advantage

(b) Teacher comparative advantage



(c) Teachers with 7+ years of experience

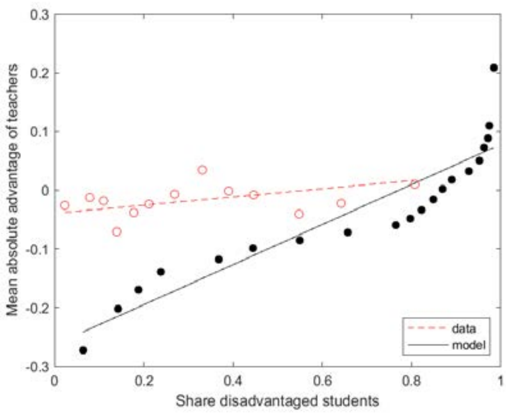
(d) Teachers that are Black



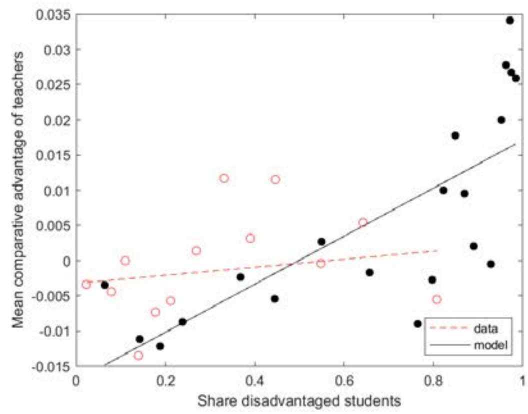
This Figure compares the allocations implied by a model in which the allocation is determined by a serial dictatorship where teachers go in descending order of their experience to the allocations we observe in the data. The data refers to all teachers in the district. The model refers to the teachers who apply in the transfer system for whom we have value-added scores.

Figure A18: Model fit: school serial dictatorship based on fraction disadvantaged (descending)

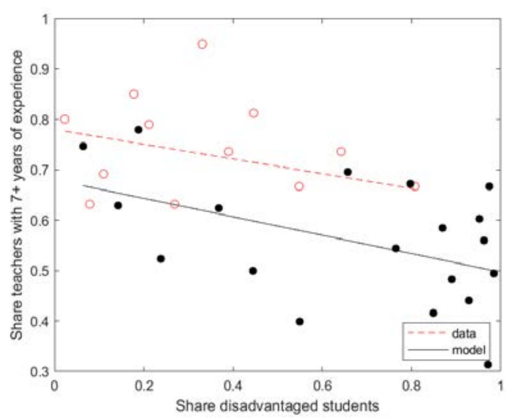
(a) Teacher absolute advantage



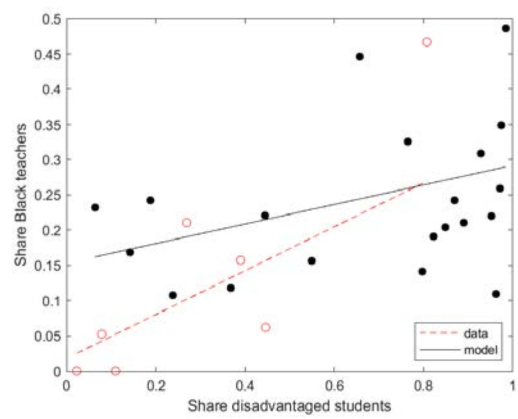
(b) Teacher comparative advantage



(c) Teachers with 7+ years of experience



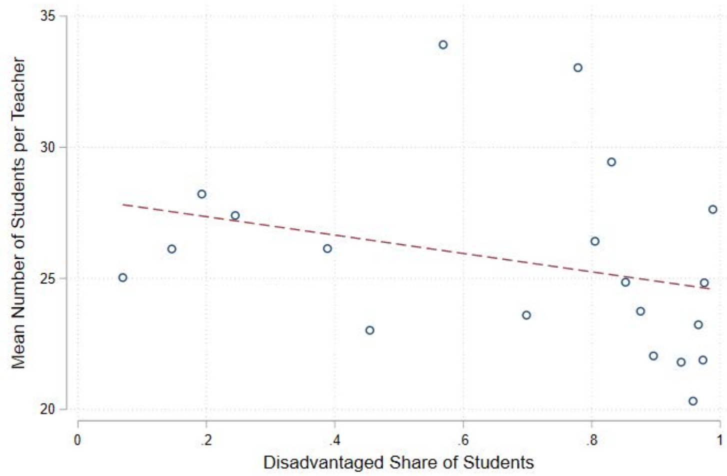
(d) Teachers that are Black



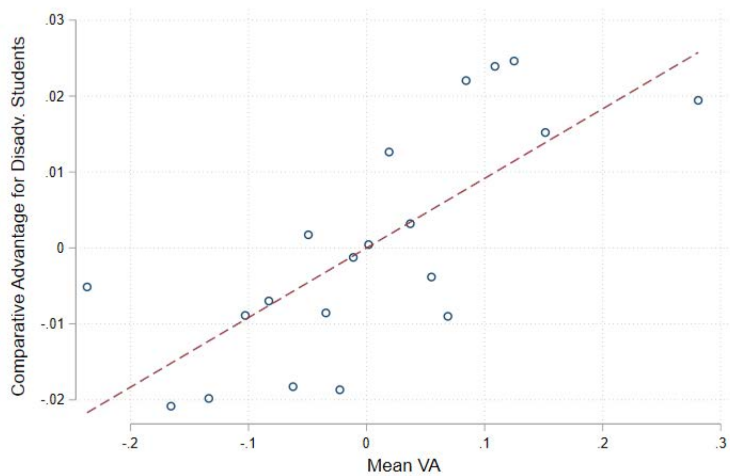
This Figure compares the allocations implied by a model in which the allocation is determined by a serial dictatorship where schools go in descending order of their fraction of disadvantaged students to the allocations we observe in the data. The data refers to all teachers in the district. The model refers to the teachers who apply in the transfer system for whom we have value-added scores.

Figure A19: Features of classes and teachers – transfer sample

(a) Class size and fraction disadvantaged

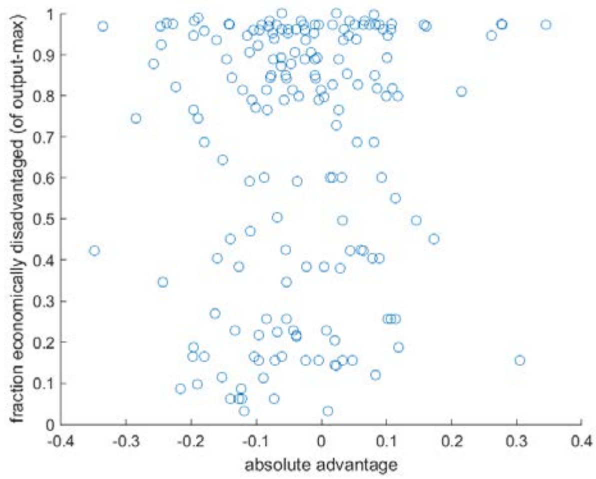


(b) Comparative advantage for disadvantaged students, and absolute advantage



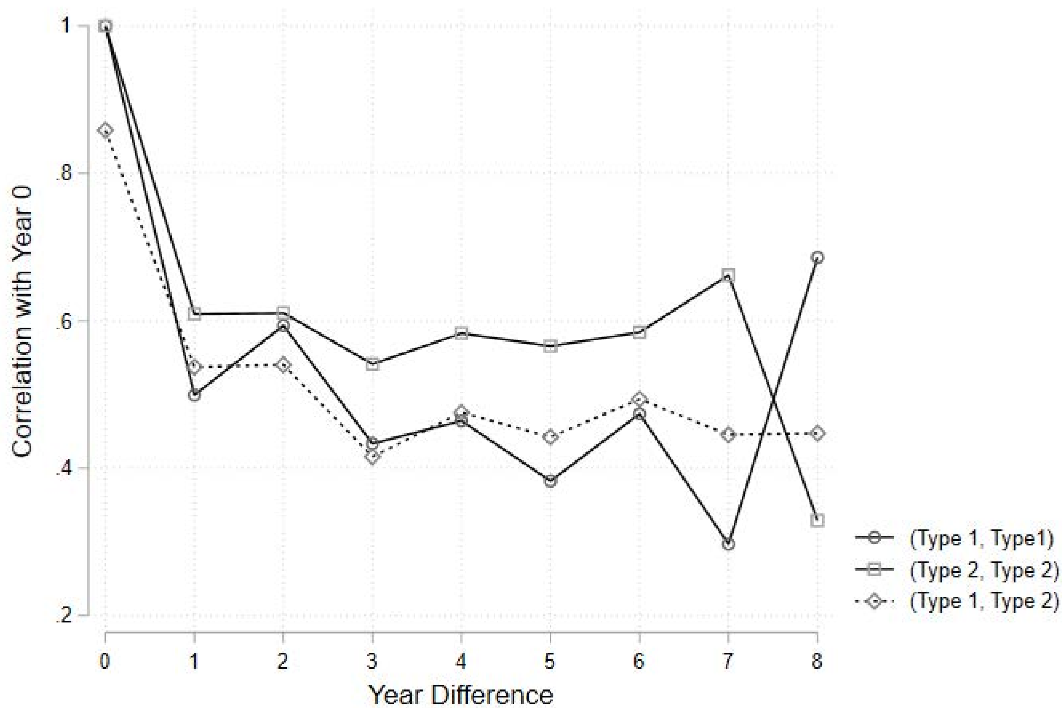
The figures show binscatters related to classroom characteristics and teacher characteristics in the transfer sample used for counterfactual analysis. The top panel shows the relationship between a school's (mean) disadvantaged share of students and a school's (mean) number of students per teacher. The bottom panel shows the relationship between a teacher's absolute advantage (x-axis) and comparative advantage in teaching economically disadvantaged students (y-axis). For this figure, absolute advantage is the average value-added across students types (rather than the value-added at a representative school) to avoid mechanical correlations between absolute and comparative advantage.

Figure A20: First-best allocation's placement of teachers, by absolute advantage



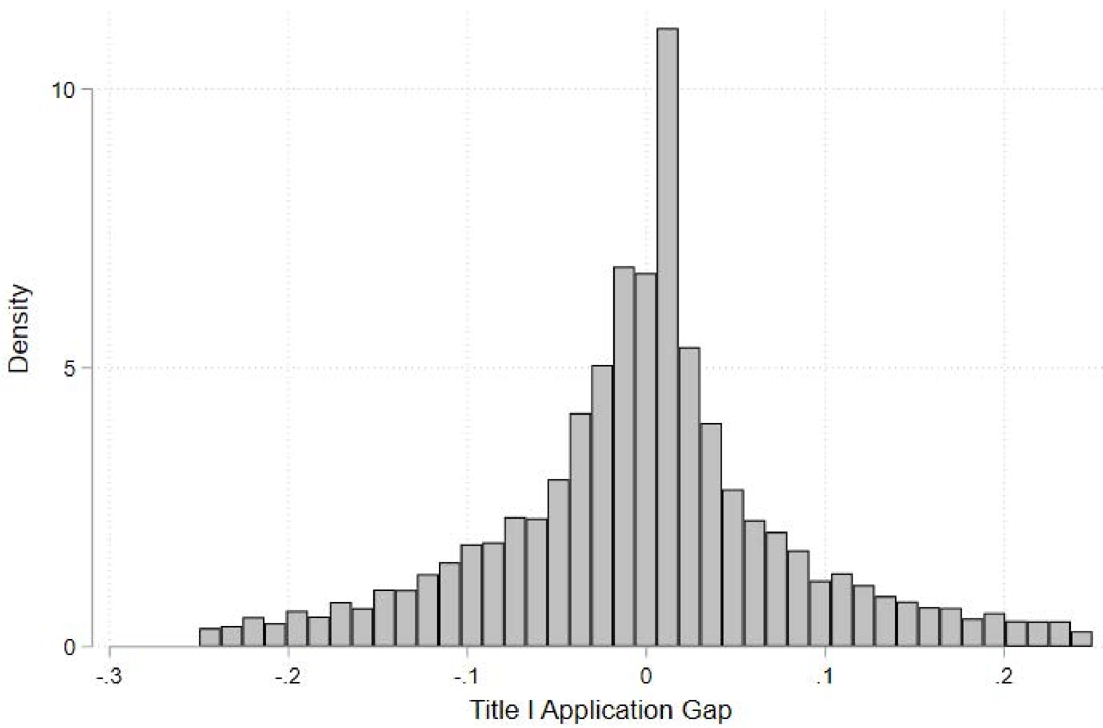
This Figure plots the first-best allocation in our transfer sample, where we divide teachers by absolute advantage and positions by fraction of students that are economically disadvantaged. Each point is an assignment of a teacher to a position.

Figure A21: Value-Added Drift Parameters



The figure shows the estimated correlations between teacher value-added in different years. The x-axis captures the year difference between the teacher's value-added measures. The three lines reflect correlations in teacher value-added within student type (1 for non-disadvantaged students, 2 for disadvantaged students) or across student type.

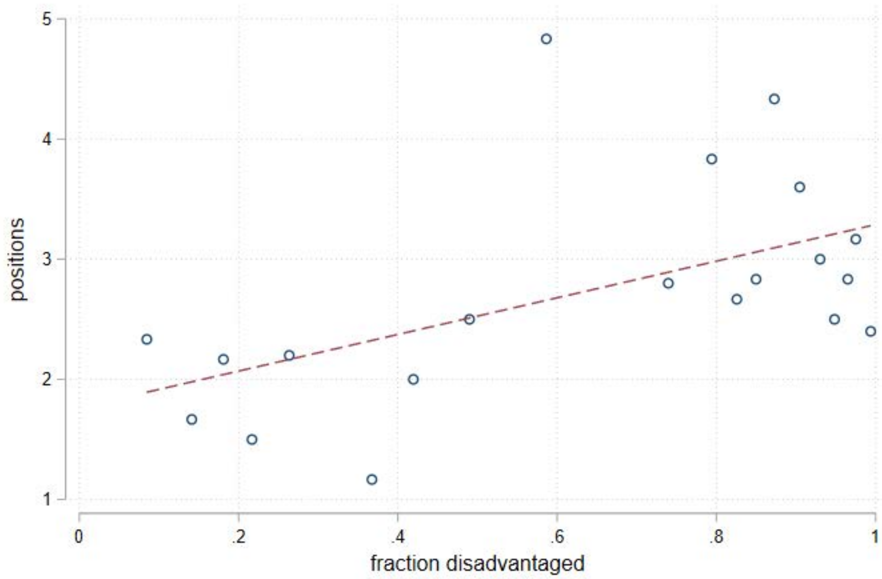
Figure A22: Title I Application Gap



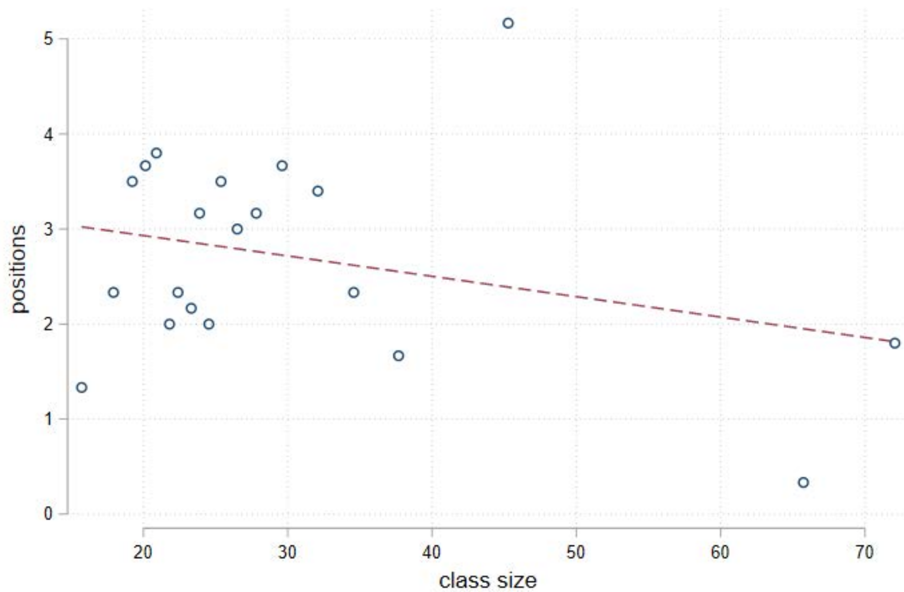
This Figure plots the distribution of the individual-level Title I application rate minus the individual-level non-Title application rate. Thus, the positive entries indicate that a teacher applies to a greater share of the Title I schools in their choice set than to the non-Title I schools.

Figure A23: Postings selection in the transfer market

(a) Positions and fraction of disadvantaged students



(b) Positions and class size



This figure shows the relationship between number of positions posted and (a) a school's fraction of students that are economically disadvantaged and (b) a school's class size. An observation is a school.