Teacher labor market policy and the theory of the second best*

Michael Bates

Michael Dinerstein

Andrew C. Johnston

Isaac Sorkin

January 2023

Abstract

The teacher labor market is a two-sided matching market where the effects of policies depend on the actions of both sides. We specify a matching model of teachers and schools that we estimate with rich data on teachers' applications and principals' ratings. Both teachers' and principals' preferences deviate from those that would maximize the achievement of economically disadvantaged students: teachers prefer schools with fewer disadvantaged students, and principals' ratings are weakly related to teacher effectiveness. In equilibrium, these two deviations combine to produce a surprisingly equitable current allocation where teacher quality is balanced across advantaged and disadvantaged students. To close academic achievement gaps, policies that address deviations on one side alone are ineffective or harmful, while policies that address both deviations could substantially increase disadvantaged students' achievement.

---

*Bates, UC-Riverside: mbates@ucr.edu; Dinerstein, University of Chicago: mdinerstein@uchicago.edu; Johnston, UC-Merced: acjohnston@ucmerced.edu; Sorkin, Stanford: sorkin@stanford.edu. We thank Peter Arcidiacono, Natalie Bau, Christina Brown, Jediphi Cabal, Dennis Epple, Felipe Goncalves, Caroline Hoxby, Peter Hull, Neale Mahoney, Richard Mansfield, Parag Pathak, Luigi Pistaferri, Camille Terrier, and numerous seminar and conference participants 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.

Inequality is a central policy concern in many matching markets. When a characteristic both confers an advantage and makes an individual more attractive to match to, the market can exacerbate inequality.[1] This mechanism might be operative in teacher labor markets, where an important policy goal is to increase the achievement of economically disadvantaged students and teachers vary widely in their effectiveness at this task (Hanushek, Kain, and Rivkin, 2004; Chetty, Friedman, and Rockoff, 2014b). Because teaching disadvantaged students presents various challenges, their schools might struggle to attract high-quality teachers. Such an unequal allocation of teachers to students may not only fail to close achievement gaps but even widen them.

Policymakers have thus argued that interventions in the teacher labor market are necessary to deliver high-quality teachers to disadvantaged students. One policy that has nearly universal take-up is switching from a system of hiring based on seniority to one of mutual consent where both the teacher and the school principal must agree to an assignment. Mutual consent has risen sharply in the US over the last few decades (Engel, Cannata, and Curran, 2017) such that 92% of large districts have now adopted it (National Council on Teacher Quality, 2022). Mutual consent fundamentally changes the economics of the teacher labor market, from a one-sided market where principals have no say to a two-sided market, where both principals and teachers must agree to a match.

In this paper, we study a large school district's two-sided teacher labor market and how it allocates teacher quality across different student types. We ask two questions: First, how effective are current policies in raising disadvantaged students' achievement? Second, what policies would most effectively raise disadvantaged students' achievement? While these questions have been the focus of large literatures, we generate new insights by focusing on the economics of two-sidedness and using novel data. As a first-best benchmark, we assume that a district maximizes disadvantaged students' achievement. The unifying theme of our findings is the "theory of the second best": when there are multiple deviations from first-best (i.e., principal and teacher preferences both deviate from those that implement the first-best allocation), these deviations can interact to produce surprisingly good allocations,

[1] Policymakers and economists have raised concerns about increasing wage inequality by worker education (e.g., Card, Heining, and Kline, 2013), the difficulty of rural hospitals attracting doctors through the US Residency Match (e.g., Agarwal, 2017), the under-representation of minority students at flagship universities (e.g., Kapor, 2020), and the increasing assortativity in the marriage market (e.g., Chiappori, 2020).

and policies that fix only one deviation can be ineffective or harmful.

Studying policy in a two-sided market is challenging for two reasons. First, we face a conceptual challenge. When both sides of a market must agree to an assignment, intuition about policies targeting one side of the market might fail for standard theory of the second best reasons. For example, the literature has argued for providing principals with more information about teacher quality (Ballou, 1996; Jacob et al., 2018a). But if all schools rank teachers according to quality, then the resulting allocation will reflect the preferences of the best teachers. If these teachers prefer advantaged schools, then inequality may rise. Second, we face an identification challenge. When both sides of a market must agree to an assignment, the equilibrium assignments reflect a combination of both sides' preferences. To infer preferences based on equilibrium assignments, researchers typically make strong assumptions that simplify the preferences on one side of the market (e.g., principals rank teachers based on quality). Inaccurate assumptions about the preferences of one side of the market may drive inaccurate policy conclusions.

We address the conceptual challenge in Section 1 by specifying a model of a two-sided matching market where allocations depend on each side's preferences and how the market clears. Teachers apply to vacancies and principals hire among applicants. We assume that equilibrium allocations are pairwise stable among teacher-vacancy pairs that are in the market at the same time (Roth and Sotomayor, 1992; Hitsch, Hortaçsu, and Ariely, 2010; Banerjee et al., 2013; Boyd et al., 2013).[2] Given the objective of maximizing the achievement of disadvantaged students, the model provides a policy benchmark: the social planner can achieve the first-best allocation if teachers prefer positions with the most disadvantaged students and principals prefer to hire the most effective teachers.

The literatures on teacher preferences and principal hiring suggest policies motivated by aligning each side's preferences with what would implement the first-best allocation. Because schools within the same district do not pay compensating differentials, the teacher literature worries that the best teachers will sort into schools with advantaged students, leaving disadvantaged students with less effective teachers (Greenberg and McCall, 1974; Antos and Rosen, 1975). The literature (e.g., Clotfelter, Ladd, and Vigdor (2011) and Goldhaber, Quince, and Theobald (2018)) thus suggests compensating teachers for taking

[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, 2022).

assignments at disadvantaged schools .[3] The principal literature worries that principals' failure to identify or hire the best teachers leads to missing productivity gains. The literature thus suggests information interventions or principal bonuses to induce principals to rank candidates by value-added (Ballou, 1996; Jacob et al., 2018a). These conclusions follow naturally from considering one side of the market—and one deviation from first-best—at a time, but do not necessarily hold in two-sided markets where both sides deviate from the first-best. Indeed, the "theory of the second best" (Lipsey and Lancaster, 1956) suggests that in the presence of multiple deviations, eliminating one may worsen outcomes. Whether these deviations are sufficiently large to overturn one-sided reasoning is an empirical question. Hence, to quantify these forces we need a credible empirical analysis.

To do so, we address the identification challenge by using detailed data, which we introduce in Section 2, from a large school district's labor market. We observe the full set of vacancies and applications, linked to student test score data. On the teacher side, we see every application. On the principal side, we see interview and offer decisions, as well as notes the principal records about applications. We observe the timing of most actions.

Data on teacher and principal actions, not just allocations, allow us to infer choice sets directly from the data—and thus to estimate preferences—under weak assumptions that we empirically motivate in Section 3. Based on institutional features and analysis of application behavior, we assume that a teacher considers all vacancies open while the teacher is active in the market and that teachers apply non-strategically to positions they prefer relative to their outside options. Based on analysis of principal behavior, we assume that a principal considers all applications when choosing whom to interview and offer. We assume that the principal ratings—rather than the decision to interview or offer—reveal principals' preferences, thus allowing the decision to interview or offer to be strategic.

In Section 4, we estimate student achievement as a function of teacher assignments. We specify a model of teacher value-added where teachers may be differentially effective with students who are economically advantaged or disadvantaged (Condie, Lefgren, and Sims, 2014; Delgado, 2022). We find some comparative advantage but that teachers mostly do not sort to students based on it.

In Section 5, we estimate teacher preferences. While a large literature has focused on

[3] Bobba et al. (2021) and Silhol and Wilner (2022) study this in one-sided teacher labor markets.

estimating teachers' preferences over schools or school characteristics (Barbieri, Rossetti, and Sestito, 2011; Engel, Jacob, and Curran, 2014; Bonhomme, Jolivet, and Leuven, 2016; Fox, 2016; Johnston, 2021), our setting and data have three advantages. First, we infer preferences from teachers' actual choices, rather than survey responses (Johnston, 2021). Second, we observe teachers' choice sets and thus separate preferences from constraints. Third, teachers take many actions, which allows for rich preference heterogeneity. Our estimates show that teachers prefer schools closer to their homes but not schools where they would have comparative advantage in raising student test scores. As others have found (Barbieri, Rossetti, and Sestito, 2011; Engel, Jacob, and Curran, 2014), teachers prefer schools with fewer disadvantaged students, though we estimate significant heterogeneity. Hence, teacher preferences differ from those that implement the first-best allocation.

In Section 6, we estimate principal preferences. As with teacher preferences, a large literature has studied principal hiring (Ballou, 1996; Boyd et al., 2011; Jacob et al., 2018b; Jatusripitak, 2018; Hinrichs, 2021). Our setting and data again offer several advantages: we infer preferences from actual choices; we observe principals' choice sets; and we identify preferences even if principals are strategic in extending interviews or offers. Principals value teachers based on observable characteristics like having a graduate degree and experience. This pattern is similar across schools with and without high concentrations of student poverty. Consistent with the literature (Ballou, 1996; Jacob et al., 2018a), while principals place some weight on a teacher's value-added (or characteristics correlated with value-added), principal's preferred candidate is rarely the one who is most effective in raising student test scores. Importantly, we reject the vertical preference model—where principals agree on the ordering of teachers—that researchers typically assume when not observing choice sets directly (see Diamond and Agarwal, 2017). Thus, principal preferences differ from those that implement the first-best allocation.

To quantify whether the deviations of teachers' and principals' preferences from those that implement the first-best allocations are large enough to overturn one-sided reasoning, we combine the estimates with the matching model. We use the estimated model to assess the current allocation and counterfactual policies. In Section 7, we show that teacher quality in the current allocation is balanced across economically advantaged and disadvantaged students, in terms of both achievement and a measure of behavioral value-added.[4] We confirm that this result reflects multiple deviations combining to produce favorable allocations. If instead of ranking teachers according to estimated preferences, principals were to only place weight on output, then we estimate that the allocation would be inequitable, as average teacher preferences suggest; thus, such policies would be harmful. By rarely ranking the most effective teacher first, principals "push back" on teacher preferences to generate parity. Consistent with second-best reasoning, the multiple deviations from first-best interact to generate a better allocation than might be expected by considering one side of the market at a time.

While the current allocation is more favorable than the one-sided literature might have expected, it does not achieve the first-best allocation. We find that reallocation can close a fifteenth of the achievement gap each year while raising average achievement in the district by 1.4 percent of a student standard deviation. In Section 8, we study the effects of partial implementations of policies that address these two deviations from the planner's solution. Specifically, we look at the effectiveness of teacher-side policies and how the effects vary with principal-side policies.

We start with teacher bonuses that only shift teachers' preferences. Bonuses for teaching economically disadvantaged students leave achievement gaps unchanged. While the bonuses shift the set of teachers who apply to schools, principals are unlikely to hire better teachers from this larger pool. Thus, the one-sided logic from the teacher preference literature suggests an ineffective policy.

We then study the effectiveness of teacher bonuses if principals hire according to value-added (implemented by some combination of principal bonuses and information interventions). Echoing the findings on the current allocation above, for small teacher bonuses, the achievement gap widens substantially because the most effective teachers sort to advantaged schools. But for large teacher bonuses, the most effective teachers now seek to teach at disadvantaged schools, and principals select them. As suggested by the planner's solution, joint teacher and principal bonuses improve disadvantaged students' outcomes.

[4] This finding is consistent with results in Sass et al. (2012), Chetty, Friedman, and Rockoff (2014b), Mansfield (2015), and Isenberg et al. (2022). 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. (Forthcoming) find similar school value-added for advantaged and disadvantaged students.

In summary, the unifying theme of the paper is the importance of the theory of the second best in analyzing a two-sided market. Subsidizing one side of the market at a time can be ineffective or harmful, even when subsidizing both sides is beneficial, and the current allocation is balanced even though teachers' preferences suggest it would not be. Reaching these conclusions requires rich data on the actions of both sides of the market.[5]

This paper fits in a growing literature on equilibrium models of the teacher labor market. These papers tend to fall into two camps. In the first, which are often outside of the US, the hiring side of the market faces constraints imposed by the government (e.g., they must hire the most experienced applicant) such that the market is essentially one-sided (Bobba et al., 2021; Combe et al., 2022; Elacqua et al., 2021; Tincani, 2021; Combe, Tercieux, and Terrier, 2022).[6] We instead focus on two-sided labor markets, which characterize nearly all teacher labor markets in the US and the hiring of permanent teachers in many non-US settings. In the second camp, several papers study two-sided markets but infer preferences from data on equilibrium allocations (Boyd et al., 2013; Bates, 2020; Biasi, Fu, and Stromme, 2021). We instead observe the actions of each side of the market, which allows us to relax the strong assumption necessary for identification in the absence of such data. We show that these assumptions deliver misleading conclusions about the relationship between teacher quality and student disadvantage in equilibrium as well as the desirability of commonly-suggested policies. Like us, Davis (2022) and Ederer (2022) study two-sided markets with data on each side's actions.[7] Unlike these papers, we estimate teacher quality based on student test scores instead of relying on observable teacher characteristics. We find that restricting teacher quality to vary only with observable characteristics changes the assessment of equilibrium and policy conclusions; for example, we find that the allocation is not balanced across advantaged and disadvantaged students in terms of teacher observables despite parity on multiple direct measures of effectiveness.

Our study carries important lessons for the analysis of labor markets. Much of the labor literature, on topics such as wage inequality (e.g., Card et al., 2018) and amenities (e.g., Sorkin, 2018), relies on matched employer-employee data where researchers only observe equilibrium allocations. These markets are two-sided, which forces researchers to

[5]Bates et al. (2022) show that these conclusions hold even if the social planner values total achievement.
[6]Bau (2022) studies an equilibrium model of school competition with school-student match effects.
[7]In work-in-progress, Laverde et al. (2021) study a two-sided market with data on each side's actions.

rely on the same identifying assumptions that led to misleading conclusions in the teachers literature. Our findings thus reinforce Oyer and Schaefer (2011)'s call for labor economists to study how firms hire workers and Card et al. (2018)'s suggestion that the labor literature on imperfect competition would benefit from "IO-style" case studies of particular markets.

1 An equilibrium model of the teacher labor market

Here, we write down an equilibrium model of the within-district teacher labor market. The model clarifies the set of factors shaping the equilibrium, allows us to define the first-best allocation, and explains when the decentralized equilibrium attains the first-best allocation.

1.1 Set-up

Teacher j derives utility u_{jk} from teaching at school k. School k's principal derives utility, v_{jk}, from hiring teacher j. Utility is non-transferable, as wages are set by the district and do not vary across assignments for a given teacher.[8]

A teacher-school assignment produces value-added VA_{jk}. Because we are interested in the achievement of disadvantaged and advantaged students, we allow the value-added to depend on the student type. Specifically, let μ_{jm} be teacher j's value-added with students of type m, where $m \in \{0,1\}$ indicates whether a student is disadvantaged. Let n_{km} be the number of students in school k of type m. Then:

$$VA_{jk} = n_{k0}\mu_{j0} + n_{k1}\mu_{j1}. \tag{1}$$

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): $\phi : \mathcal{J} \rightarrow \mathcal{K}$ so that $\phi(j) = k$, the school k to which teacher j is assigned.[9] Denote by Φ the set of all possible assignments.

[8]This assumption also excludes transferable non-pecuniary benefits, such as favorable class assignments. In Appendix B, we test for evidence of favorable class assignments by experience level and fail to reject no relationship.

[9]For simplicity, we specify each school as having a single position. When we estimate the empirical model, schools may have multiple positions.

1.2 First-best allocation

We are interested in policies that increase the achievement of disadvantaged students. We take as given the set of teachers and positions the district has and ask how to assign them. 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.[10]

The district values the achievement of disadvantaged students:[11]

$$\max_{\phi \in \Phi} \left\{ \sum_{j \in \mathcal{J}} n_{k1} \mu_{j1} \right\}. \tag{2}$$

The structure of the first-best allocation is simple: rank teachers in descending order by value-added with disadvantaged students and rank classrooms in descending order by the number of disadvantaged students. Then assign the strongest teacher to the classroom with the largest number of disadvantaged students and so on.[12]

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.

1.3 Decentralized equilibrium

Our equilibrium concept is (timing-constrained) pair-wise stability. Schools meet with all teachers who are in the market at the same time. 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

[10] 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.

[11] In Bates et al. (2022), we include advantaged students' achievement and teacher utility.

[12] Table 4 (Part 1) shows that our results are very similar if we hold class sizes constant.

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). We use DA in order to find stable equilibria, not because DA is actually used in this market.

When does the decentralized equilibrium correspond to the planner problem? Suppose that teachers rank schools according to the number of disadvantaged students ($u_{jk} \propto n_{k1} \forall j,k$) and principals rank teachers according to total output ($v_{jk} \propto VA_{jk} \forall j,k$). Then in the absence of comparative advantage or timing restrictions, the decentralized equilibrium—which is unique in this case—corresponds to the planner's solution. Notably, this combination of rankings is what the *joint* implementation of hard-to-staff school bonuses and guided principal hiring would achieve. Of course, the theory of the second best says that aligning only the principal or the teacher with the planner may not improve outcomes.

Empirically, we are then interested in the extent to which teacher and principal preferences align with those that decentralize the planner's solution. We are also interested in whether the other factors we have abstracted from—timing and comparative advantage—affect the gap between the decentralized equilibrium and the planner's solution.

2 Data and institutional context

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 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 3. Second, the file records the city, zip code, and 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 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 platform to the state's administrative data, using names, birth dates, and the last four digits of teachers' social security numbers. 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.

2.3 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, where each school posts openings. 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 what Papay and Kraft (2016) find, 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 necessitates 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 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 but little marginal cost to apply to an additional posting.

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.

Eligibility: Teachers are eligible for positions if they have the necessary certification. We will focus on the market for elementary-school classroom teachers because the common certification allows us to reliably classify which teachers are eligible. We can also infer elementary school teachers' quality from systematic gains in their students' test scores because teachers in these positions are typically responsible for instruction in the tested subjects.

3 The vacancy posting, application, and hiring process

In this section, we describe our model of teacher and principal actions in the labor market. We specify our model assumptions, consider how violations of the assumptions might manifest in the data, and show empirical evidence consistent with the assumptions. Our empirical analysis will also include robustness checks around possible alternate assumptions. We defer a discussion of the pair-wise stability assumption until Section 7.

3.1 The teacher perspective

3.1.1 How we model applications

The district's labor market consists of potential teachers, indexed by j, and a 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.[13] 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'. These positions comprise her choice set, $\mathcal{P}_{jt}$. There is no marginal cost to applying and there is no limit on the number of applications she can submit within the choice set. Let a_{jpt} be an indicator for whether teacher j applied to position p in year t. A teacher's application a_{jpt} is 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.[14] A teacher submits an application to position p if:

$$a_{jpt} = \mathbf{1}\{u_{jpt} > \max\{u_{jct}, u_{j0t}\}\}, \tag{3}$$

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

3.1.2 Model assumptions

There are three key assumptions that underlie this model of teacher application behavior. First, applications are non-strategic: if a position is more appealing than the outside option and current position, then the teacher applies. Second, the teacher considers all vacancies that overlap with her timing. Third, the set of positions the teacher sees is exogenous.

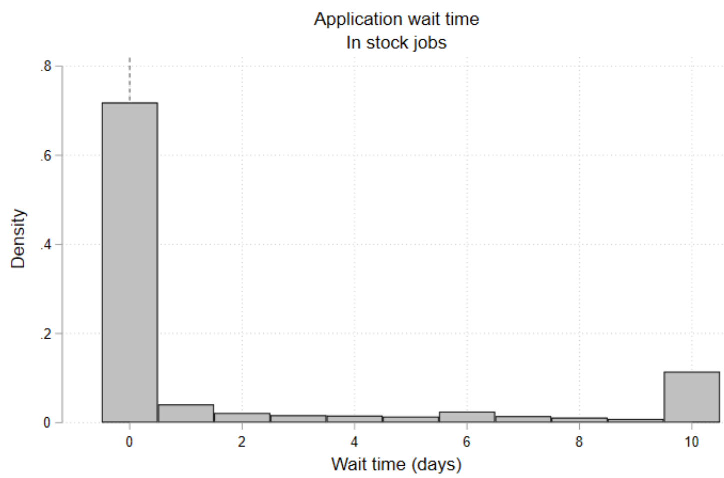
First, assuming nonstrategic applications is reasonable because of two institutional features and one data analysis. First, the marginal cost of applying to a vacancy is effectively zero (it just requires clicking submit given already uploaded materials) so it is reasonable that a teacher just compares a given position to the outside option. Second, principals do not see the teacher's other applications, which limits complicated signaling stories. Third, if teachers were instead strategic in submitting applications, then most models would imply a dynamic portfolio strategy where teachers might delay when they apply to a vacancy. We

[13] We assume entry into the system is exogenous. We discuss selection into the system in Appendix C.

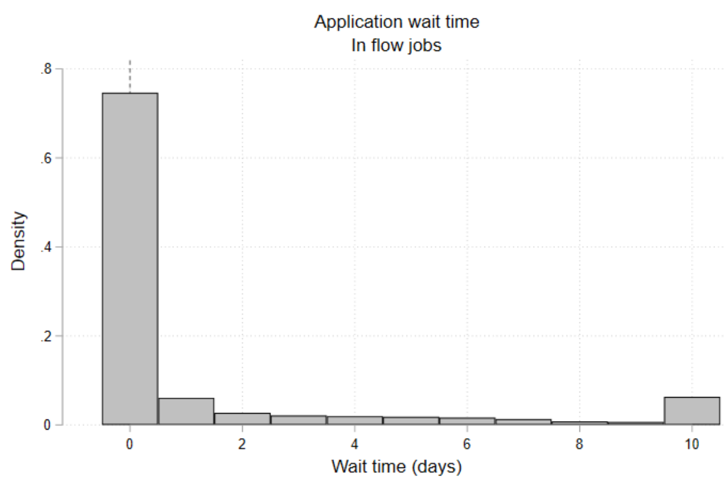
[14] We assume that any post-application steps necessary to be assigned to a position – e.g., interviews – are costless. In our data, teachers with multiple interviews are so rare that even if interviews are costly, they are rare enough that it is unlikely teachers consider dependence across applications.

Figure 1: 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. The final category corresponds to waiting at least 10 days. The median wait time is zero in both figures.

investigate this empirically by constructing a measure of a teacher's wait time to apply to a vacancy. We calculate 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.[15] The top panel of Figure 1 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. 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. We thus find minimal waiting to apply to positions, such that teachers are unlikely to be engaging in dynamic portfolio strategies.

Second, it is reasonable to assume teachers consider all vacancies because of the same evidence. If teachers were unaware of some open vacancies, then we would expect teachers to apply frequently after the first opportunity to do so. We see little evidence of such delayed applying. This pattern could reflect teachers missing a vacancy when it is posted and never search for older vacancies. But we see the opposite – on the first day of applying, teachers apply to old and new vacancies, with a mean vacancy length of 23 days (Appendix Table A2, panel B).

We construct a teacher's start (r) and end (r') (search) date as the dates of her first and last application, respectively. We thus estimate fairly large choice sets out of which teachers make a large number of choices, which helps us estimate preference heterogeneity. Specifically, the mean choice set size is 159 (median: 139), and the mean number of applications is 23 (median: 8).

Third, it is reasonable to assume the set of positions the teacher sees is exogenous for three reasons. First, it is hard to predict when teachers enter the system. We look at teacher value-added and find that above and below median teachers apply at similar times (see Appendix Table A3b). Second, it is hard to predict when teachers exit the system. Many teachers—including those who do not successfully transfer—stop applying long before the end of the hiring season (9% in April or before, 16% in May, 22% in June; see Appendix Table A2, panel C). This pattern suggests that exit is likely driven by shocks unrelated to

[15] Let $\mathcal{A}_{jt}$ denote the set of days where teacher j applied to at least one vacancy in year t, with $a_{jt} \in \mathcal{A}_{jt}$ in days. Let b_{kt} be the 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: $w_{jkt} \equiv c_{jkt} - \min_{a_{jt} \in \mathcal{A}_{jt}: a_{jt} \geq b_{kt}} a_{jt}$.

accepting a job or to the nature of the jobs being posted. Third, it is hard to predict when positions are posted. For example, even within school, there is vast variation in the timing of postings across years: pooling across the years in our data, 89% of schools that post jobs in July also post jobs in April, and a similar pattern holds for schools with April postings (see Appendix Table A3c). Combined, these three features suggest that there is likely little correlation between teacher characteristics and the set of vacancies that they see.[16]

3.2 The principal perspective

3.2.1 How we model principal behavior

Each position p is associated with a principal with the same index. Principal p derives utility v_{jpt} from teacher j holding the position in year t. We model a principal as giving teacher j a positive rating ($b_{jpt} = 1$) if the utility is positive: $v_{jpt} > 0$. A positive rating is at least one positive outcome: recording a positive note about the application, offering an interview, or extending a job offer.

3.2.2 Model assumptions

There are two assumptions underlying our model of principal behavior. First, principals prefer applicants who receive a positive outcome to those who do not. Second, principals consider all applicants.

The note-taking system is supportive of the first assumption. Principals may be strategic in deciding on interviews or offers if such actions are costly and a preferred teacher may have a low probability of accepting. Because the note-taking system allows principals to rate applicants with no direct consequences, principals can reveal their preferences while remaining strategic in consequential actions.[17]

The second assumption is reasonable because we see no relationship between when an applicant applied and the applicant's outcome. The applications that receive ratings are

[16]Table 4 (Part 2) shows robustness to a seven-day buffer on both ends or to dropping teachers who only apply to one school. If choice sets are restricted, then fixing the deviations is further from first-best.

[17]While our assumptions allow for strategic interviews and offers, we do not find evidence that strategic behavior is common enough to affect our conclusions. Table 4 (Part 3) shows that results are robust to instead modeling principal behavior with a rank-order logit, including where we restrict to only active choices (i.e., drop applications with no records in the note-taking system).

similar in timing to those that the principals do not rate (see Appendix Table A4).[18]

4 Production of student achievement

In this section, we first lay out 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.

4.1 Model

Given our interest in outcomes for disadvantaged students, we allow teacher value-added to differ between advantaged and disadvantaged students.[19] This choice follows the quickly expanding literature documenting match effects or allowing for comparative advantage (Dee, 2004, 2005; Condie, Lefgren, and Sims, 2014; Jackson, 2013; Aucejo et al., 2022; Delgado, 2022; Graham et al., 2020; Biasi, Fu, and Stromme, 2021; Bau, 2022).

We use notation that follows Chetty, Friedman, and Rockoff (2014a) and Delgado (2022). 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

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

[18]Table 4 (Part 4) shows that results are robust to varying which applicants we assume principals consider.

[19]In robustness checks in Table 4 (Part 5), we consider two alternative splits of students: race and lagged student achievement. We find that our substantive conclusions are nearly identical.

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 (2014a), 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 standard assumptions to identify the model (see Appendix D).

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 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.

4.2 Estimation

We estimate our model in three steps using math scores and data from the whole state.[20] In the first step, we estimate the coefficients on student 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 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. Using shrunken estimates and prior data implies that we use

[20]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 (2014a) for evidence). In Section 7 we show robustness to including a teacher's value-added on behavioral outcomes.

the information available to policy-makers. See Appendix D.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 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 (2014b)). Third, we include teacher-year fixed effects in the residualization step, rather than teacher-class-student type effects as in our baseline. See Appendix D.3 for details.[21]

4.3 Validation of the match effects model

To validate our value-added model, we use a version of Chetty, Friedman, and Rockoff (2014a)'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 A5 shows that the math value-added estimate is an unbiased predictor of residualized output, with a tight confidence interval around 1.05. Figure A1 shows that forecast unbiasedness holds throughout the distribution of teacher value-added. Column (4) of Table A5 shows mean forecast unbiasedness nearly holds for transferring teachers while the last two columns show mean forecast unbiasedness even for cases where teachers switch between classrooms with very different compositions or sizes.

We conduct a similar test for the comparative advantage component of value-added. 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 A2 shows that forecast unbiasedness holds throughout the distribution. Appendix D.4 further assesses the validity of the comparative advantage component of value-added, providing inference around relevant structural parameters, likelihood tests, and additional validation around transferring teachers.

[21] Table 4 (Part 6) shows that our results do not depend on which value-added model we use.

5 Teacher preferences

5.1 Parameterization

We adopt a characteristics-based representation of teacher utilities over positions, which helps us to estimate preference heterogeneity. Teacher utilities over positions are:

$$u_{jpt} = -\gamma d_{jpt} + \pi_j VA_{jpt} + \beta_j X_{pt} + \eta_{jt} + \varepsilon_{jpt}. \tag{7}$$

Teacher utility for the outside option is $u_{j0t} = \varepsilon_{j0t}$. d_{jpt} is the one-way commute time (in minutes) between the teacher and the position and will serve as a numeraire for exposition. 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: $AA_{jt} = n_{0t}\hat{\mu}_{j0t} + n_{1t}\hat{\mu}_{j1t}$, 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, the coefficient on VA_{jpt}, π_j, captures the strength of teachers' preferences for schools where their comparative advantage is high. We allow for preference heterogeneity by including a random coefficient in π_j:

$$\pi_j = \bar{\pi} + \sigma^{VA} \nu_j^{VA}, \tag{8}$$

where $\nu_j^{VA} \sim^{iid} N(0,1)$.

X_{pt} is a vector of observed characteristics of positions: the fraction of a school's students that are (1) economically disadvantaged (e), (2) above the median in prior year math test scores (s), (3) Black (b), and (4) Hispanic (h). We allow for heterogeneous preferences:

$$\begin{aligned} \beta_j^e &= \beta_{j0}^e + \beta_{j1}^e AA_{jt} + \sigma^e \nu_{jt}^e \\ \beta_j^b &= \beta_{j0}^b + \beta_{j1}^b AA_{jt} + \beta_{j2}^b Black_j + \sigma^b \nu_{jt}^b \end{aligned} \tag{9}$$

where $Black_j$ is an indicator for teacher race category and ν is a vector of independent, standard normal random coefficients, which captures the standard deviation of idiosyncratic preferences. The equations for lagged achievement and Hispanic are parallel.[22]

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:

$$\eta_{jt} = \lambda Z_{jt} + \rho CM_{jt} + \sigma^{\eta}\nu^{\eta}_{jt}. \tag{10}$$

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. These variables also soak up most of the salary variation across teachers. 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). Through CM_{jt}, we allow unobserved heterogeneity to be correlated with CA_{jpt} and X_{pt}. Finally, ν^{η}_{jt} is an independent standard normal random effect.[23]

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:

$$Pr(a_{jpt} = 1) = \frac{exp(V_{jpt})}{1 + exp(V_{jct}) + exp(V_{jpt})} \text{ and } Pr(a_{jpt} = 1) = \frac{exp(V_{jpt})}{1 + exp(V_{jpt})}, \tag{11}$$

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

5.2 Estimation and Identification

The data we use to estimate teacher preferences are applications to positions, and the method we use is maximum simulated likelihood, where we simulate from the normal distributions of the random coefficients. Let n index each simulation iteration and let $A_{jptn}(\theta)$

[22] Table 4 (Part 7) shows that our results are robust to allowing for correlation in the random coefficients.

[23] In Table 4 (Part 8), we consider binary logits, and show that our results are robust to either omitting random effects, or to including various combinations of teacher and school random and fixed effects.

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:

$$L_{jt} = \frac{1}{100} \sum_{n=1}^{100} \prod_{p \in \mathcal{P}_{jt}} (a_{jpt} A_{jptn}(\theta) + (1 - a_{jpt})(1 - A_{jptn}(\theta))), \tag{12}$$

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

$$l = \frac{1}{J} \sum_{j} log L_{jt}. \tag{13}$$

In Section 3 we argued that the institutions and data are consistent with teachers applying 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. Similar reasoning applies for mean coefficients, and observed and unobserved preference heterogeneity.

We seek to predict teachers' valuations over positions rather than causal effects of changes in characteristics on choices. In counterfactuals, we give utility bonuses as a function of school characteristics and so do not assume that teachers value money or these characteristics. 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. We do not rely on having consistently estimated the causal effect of commute time, however, because we only make relative comparisons of the costs of various policies.

5.3 Teacher Preference Estimates

Table 1 presents the teacher preference estimates. First, teachers prefer positions with more 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 unob-

Table 1: Teacher preference estimates

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

servable heterogeneity. For example, teachers with higher absolute advantage have relatively lower preferences for schools with more disdvantaged students. We also find a large positive same-race premium for Black teachers and schools with large fractions of Black students. In terms of unobserved heterogeneity, we typically find substantial dispersion in the random coefficients. For example, a standard deviation of the random coefficients on fraction disadvantaged is 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 2 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. 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 of percentiles. The figures also emphasize that there is substantial heterogeneity in teachers' rankings of positions: across the support of these characteristics, the range from the 10th percentile in the teacher distribution to the 90th is very large.

Hence, not only do teacher preferences deviate from those that would decentralize the planner's solution, they are negatively correlated. With minimal assumptions and data on real choices, we confirm the findings of the teacher preference literature regarding mean preferences but estimate considerable heterogeneity.

6 Principal behavior

6.1 Parameterization and identification

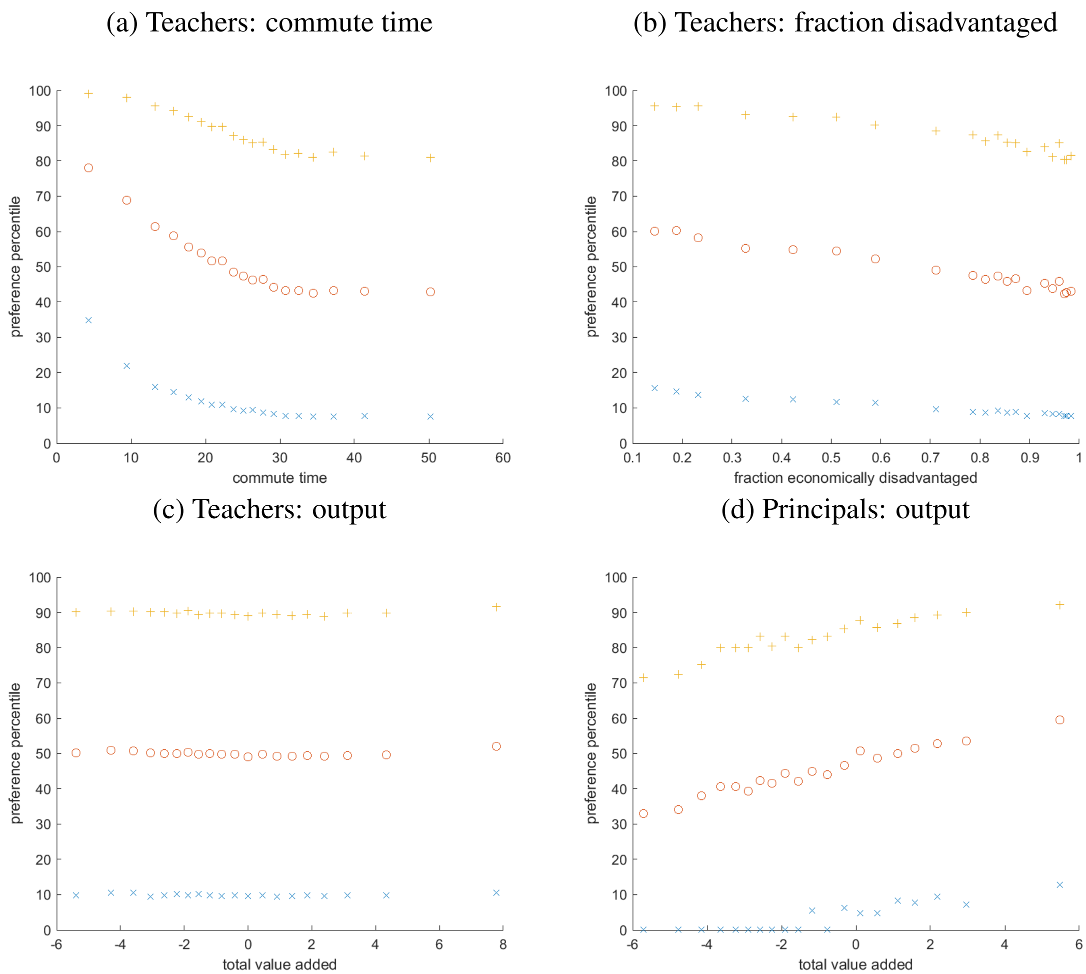
We adopt a characteristics-based model and parameterize v_{jpt} to be a linear function of position and teacher characteristics, a random effect, and an idiosyncratic teacher-position error:

$$v_{jpt} = \alpha_p W_{jpt} + \sigma_\kappa \kappa_{pt} + \upsilon_{jpt}. \tag{14}$$

To allow principal behavior to possibly align with output, W_{jpt} includes j's total value-added at school $k(p)$. We further include 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.[24] 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 exclude salary because principals in

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

Figure 2: 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 the 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 the 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 pointwise 10th and 90th percentiles, respectively.

our empirical context do not have to pay teacher salaries out of a school budget. 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 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.

Table 2: Principal valuation estimates

	Estimate	Standard Error
Characteristics		
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

	Estimate	Standard Error
Black	-1.035	0.227
Black x Title I	1.722	0.453
Black x Frac. Black	0.396	0.267
Black x Frac. Black x Title I	-0.253	0.511
Hispanic	-0.690	0.454
Hispanic x Title I	0.450	0.561
Hispanic x Frac. Hispanic	2.259	2.219
Hispanic x Frac. 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 valuation 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.

As with teachers, identification is straightforward given our characterization of the process in Section 3. We observe the set of applications that a principal receives and we observe whether a principal gives an application a positive outcome. 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 valuations 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 assumption still holds. 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 (see Table 4 (Part 3)).

6.2 Estimates

Before presenting estimates from our baseline model, we consider what types of characteristics determine principal ratings. Appendix Table A6 presents the changes in pseudo-R^2s from including different sets of observable teacher characteristics. The main set of characteristics that explain ratings decisions are various observable characteristics of teachers: experience, licensing, certification, and Praxis scores.[25] While one might think that these characteristics would predict value-added, in Appendix Table A7 we show that they have very limited predictive power. Indeed, value-added by itself or in addition to other characteristics generates very small changes in model fit.[26]

Despite the small explanatory power of value-added in principal decisions, Table 2 shows that principals do favor teachers with higher value-added in our baseline model.[27] We also observe significant heterogeneity, as Title I school principals rate Black and Hispanic teachers more positively than non-Title I principals do. To help interpret the strength of the value-added relationship, Panel (d) of Figure 2 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. Consistent with the idea that observed characteristics poorly predict value-added, Appendix Figure A3 shows that if we omit value-added from the principal model then the relationship dramatically flattens.

Hence, principal valuations deviate from those that would implement the planner's solution, as principals rank teachers based on predictable and unpredictable factors not related to value-added. Whether the positive relationship between rankings and value-added is strong enough to generate allocations close to the planner's solution depends on how both sides combine in equilibrium.

7 Understanding the current allocation

We combine the estimated market timing from Section 3, the estimated match-specific output from Section 4, the estimated teacher preferences from Section 5, and the estimated principal valuations from Section 6 to simulate the market equilibrium.

[25] In Table 4 (Part 9), we show that whether we include licensing, certification, and Praxis scores in the principal model has little effect on our results.

[26] EVAAS, the state of North Carolina's value-added measure, has even less explanatory power. As principals have access to this information, it is unlikely that the estimated weights principals place on value-added are due to measurement error in our estimates of value-added. Our results are quantitatively robust to significant amounts of attenuation. See Table 4 (Part 10).

[27] See Appendix E for the likelihood, which closely parallels the one for teachers.

7.1 Simulation details

We consider allocating the set of *teachers who apply for positions* in the district in the 2015-2016 cycle, including teachers who are not currently in the district. 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: we end up with 178 elementary school teachers and 296 positions. To avoid the possibility of artificial imbalance playing a role in our estimates (see Ashlagi, Kanoria, and Leshno (2017)), in each of 200 simulation runs we randomly drop positions so that there are the same number of teachers and positions.[28]

While we estimate a distribution of random coefficients, in simulations we use the single draw of the random coefficients per teacher and principal that maximizes the likelihood for the teacher or principal. We draw i.i.d. type I errors.

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.

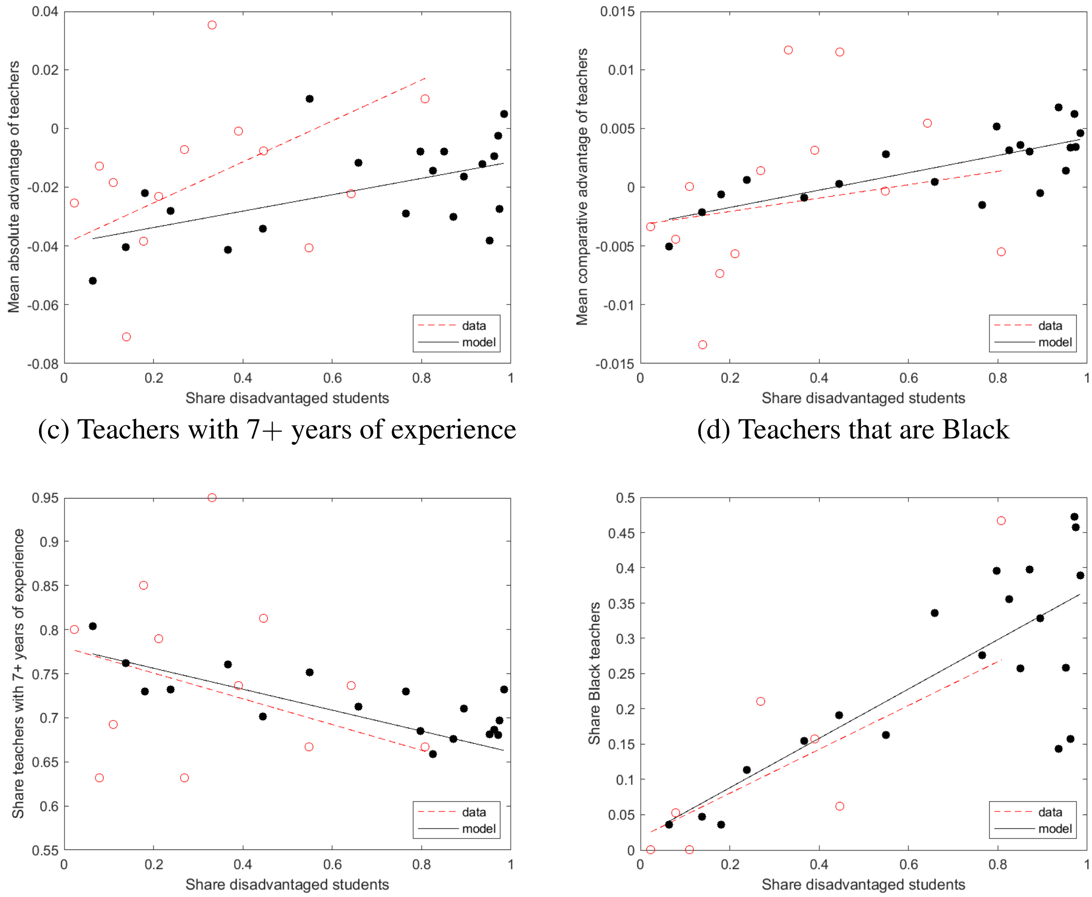
For teachers and vacancies that are not in each other's choice sets, we assign a large negative number to the valuations. We do not include an outside option when we run DA. Given that we impose balanced markets, all teachers are hired and all positions are filled.

7.2 Model fit

We begin by considering the model's fit under the status quo. Because we estimate several model components fairly directly from data, fit largely highlights how well our market equilibrium assumption (pairwise stability) performs. Figure 3 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

[28] In practice, we have also explored many of our results without dropping positions and we found that our results are very similar. Results are available upon request.

Figure 3: 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. The intercepts are normalized to be equal.

teacher experience and whether teachers are Black. The model slightly underpredicts the slope in absolute advantage.[29]

Figure 4 (and Table 3) shows that in the estimated status quo, disadvantaged students are assigned slightly *better* teachers than advantaged students. This feature matches the data.

[29]We also find our model fits better than models with alternate equilibrium assumptions: a teacher serial dictatorship ordered by absolute advantage or experience or a principal serial dictatorship ordered by fraction of students that are economically disadvantaged. Results are available upon request.

Table 3: Current allocation, alternative policies, and first-best

Description	VA disadv.	VA adv.	mean VA
Status quo	-0.018	-0.038	-0.025
Status quo with school propose	-0.018	-0.038	-0.025
Satus quo with all options (timing)	-0.015	-0.031	-0.020
Status quo and teachers rank by N disadv.	-0.020	-0.028	-0.023
Status quo and principals rank by VA	-0.047	0.004	-0.030
Status quo, and previous two changes	0.033	-0.105	-0.014
First-best	0.037	-0.104	-0.011

This table displays numbers corresponding to the allocations plotted in Figure 4, as well as the overall achievement per student.

Appendix Table A8 shows that this feature holds looking at all teachers in the district (not just those in the transfer system). This pattern also holds in raw test score gains, estimates from alternate test score value-added models, and behavioral value-added estimates, and holds in districts in North Carolina outside our focal district. The Table also shows that while we allow for comparative advantage, teachers do not sort on the basis of it.

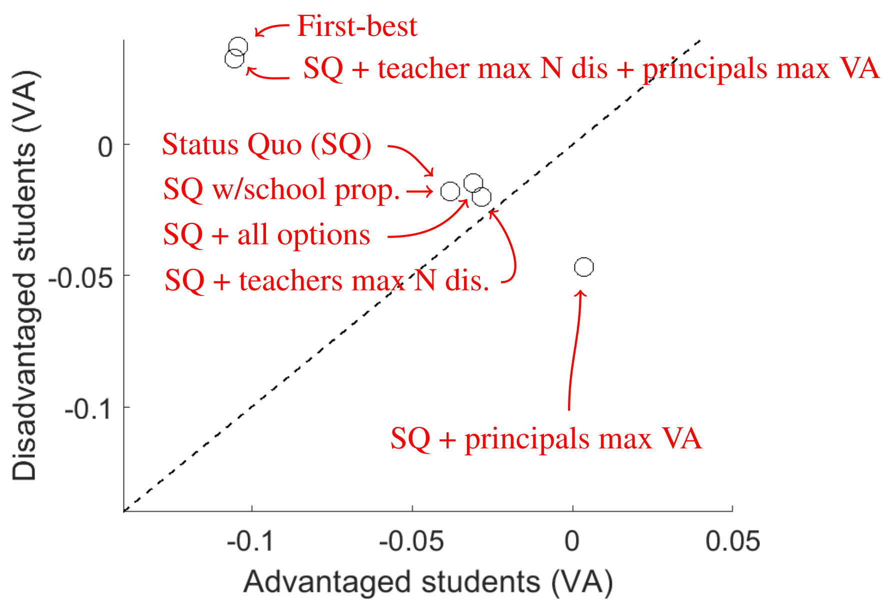
7.3 The importance of second-best reasoning

In the last subsection, we documented that advantaged students have no more effective teachers than disadvantaged students. Relative to the structure of teacher preferences, this balance is surprising in that teachers' revealed preference is strongly averse to teaching at schools with disadvantaged students. In this section, we explain this result through the economics of two-sided markets and the theory of the second best.

A couple of subtle explanations play no role in explaining the current allocation. First, there is no room for equilibrium selection. Changing the equilibrium from the teacher-proposing equilibrium to the school-proposing equilibrium has no effect on the allocation. Second, timing has little role. Changing timing so that all vacancies and teachers are active at the same time increases output by a similar, small amount for both types. We show these and other allocations in Figure 4 and Table 3.

Aligning teacher and principal preferences with the planner's solution shows that there

Figure 4: Current allocation, alternative policies, and first-best



This figure simulates the trade-off between student achievement for economically advantaged and disadvantaged students. The status quo uses teacher and principal estimated preferences and restricted choice sets, and solves for the teacher proposing stable allocation. The first-best is the allocation where the planner maximizes the achievement of disadvantaged students. The Figure plots averages over 200 simulations.

are important interactions between both sides of the market, such that thinking about one side at a time leads to ineffective or harmful policy ideas. First, if teachers had preferences that would decentralize the planner's solution—they only care about the number of disadvantaged students in a school—then the allocation is little changed. Thus, a natural teacher-side policy is ineffective. Second, if principals had preferences that would decentralize the planner's solution—they only care about the output in the match—then the allocation is worse for equity and resembles what we might expect based on the structure of teacher preferences.[30] Thus, a natural policy based on one-sided reasoning is harmful.

One-sided reasoning is misleading here because of the theory of the second best: preferences on both sides of the market deviate from the preferences that decentralize the planner's solution, but these deviations interact to generate surprisingly favorable allocations.

[30]The allocation is also worse for efficiency: per student output declines by about 0.005σ.

Were we to eliminate the deviation on the principal side of the market and have principals order teachers by value-added, then the strongest teachers would reach their most preferred schools. Given the structure of teacher preferences, this change would lead advantaged students to have much more effective teachers. Hence, by placing weight on factors other than value-added, principals "push back" on teacher preferences and overcomes differences in applicant pools across positions.

Addressing both the conceptual and identification challenges mentioned in the introduction is necessary to reach these conclusions. Specifically, we need an equilibrium model, and data to identify preferences from actions rather than equilibrium assignments. With data only on equilibrium assignments, typically one assumes that one side of the market has vertical preferences, which fills in the choice sets for the other side of the market. If we had (incorrectly) assumed principals have vertical preferences in value-added, then we would have concluded that the status quo was very unfavorable to disadvantaged students, and teacher bonus policies by themselves were effective.

Figure 4 (and Table 3) shows that there are substantial gains in the first-best. Disadvantaged students gain about 0.055σ, or about one-fifteenth of the unconditional achievement gap that we document in Appendix Table A8. While these numbers refer only to teachers in the transfer system, in Appendix Table A9 we show that these gains are similar if we look at all teachers in the district. Naturally, these gains are not costless—they come somewhat at the expense of advantaged students, whose teacher quality suffers, but total output still increases (by about 0.015σ in both the transfer and overall samples).

Finally, Figure 4 (and Table 3) shows that the combination of the two policies mentioned above—teachers rank schools based on the number of disadvantaged students and principals rank teachers based on projected output—comes close to decentralizing the first best (it achieves 92% of the first-best, the remaining gap is due to comparative advantage and timing). Thus, in Section 8 we study policies that move us closer to this point.

7.4 Different objectives

Our social planner maximizes disadvantaged students' output. Here, we consider how our results might change with alternate objectives.

First, the social planner may place weight on other forms of output, not just math test scores. We estimate teachers' value-added on an index of behavioral outcomes and find that behavioral value-added is still balanced across advantaged and disadvantaged students (Appendix Table A8).

Second, the social planner may place weight on other agents, not just disadvantaged students. First, the social planner may place equal weight on all students. We consider this objective in Bates et al. (2022) and estimate that aligning principals' preferences with the social planner's objective function still lowers total academic achievement. Aligning teachers' preferences with the social planner's, though, can lead to some total output gains. Second, the social planner may place weight on teacher utility. In Section 8, we constrain the policies we consider to make each teacher weakly better off than in the status quo.

7.5 Robustness

In Table 4, we report the robustness checks we have mentioned in footnotes throughout the text. The following three basic findings are robust across all of these alternatives: first, there are large gains from moving to the first-best; second, fixing one of the deviations from what decentralizes first best (making teachers value the number of disadvantaged students or principals maximize value-added) is either ineffective or harmful; and third, that fixing both comes very close to implementing the first-best (restricted choice sets play a quantitatively larger role when we restrict teacher choice sets to the first day).

8 Teacher bonus counterfactuals

In this section, we consider policies that may move the allocation closer to the first-best. We compare teacher bonus policies that cost the district equivalent amounts while holding all teachers harmless. We then interact these bonuses with principal-side policies.

8.1 Implementation details

The district offers a two-part bonus on the basis of a teacher-position characteristic, z_{jpt}, where each teacher receives a lump-sum amount, b_0, and a bonus b_1 per unit of character-

Table 4: Robustness: disadvantaged achievement

	Status quo	All Options	Principal Max VA	Teach Max Frac Dis	Previous Two	First Best
Baseline	-0.018	-0.015	-0.047	-0.020	0.033	0.037
1. Hold class sizes constant: baseline uses class size						
Constant class size	-0.025	-0.027	-0.051	-0.026	0.018	0.020
2. Vary choice set construction for teachers						
7 day buffer	-0.019	-0.017	-0.053	-0.018	0.034	0.037
First day choice sets only	-0.023	-0.015	-0.040	-0.022	0.016	0.037
Drop single app. teachers	-0.018	-0.017	-0.047	-0.020	0.031	0.035
3. Estimate principal preferences using rank order logit: baseline is binary logit						
All data	-0.018	-0.015	-0.046	-0.025	0.032	0.037
Active choices	-0.018	-0.016	-0.047	-0.031	0.032	0.037
Hire outcome only	-0.017	-0.014	-0.047	-0.021	0.033	0.037
4. Vary window in which we estimate principal preferences: baseline is all applications						
W/in 2 weeks of hire	-0.017	-0.014	-0.047	-0.021	0.033	0.037
First half	-0.0173	-0.014	-0.047	-0.020	0.033	0.037
Second half	-0.0175	-0.016	-0.047	-0.026	0.033	0.037
5. Vary student type split: baseline is economic disadvantage						
Achievement	-0.016	-0.014	-0.047	-0.020	0.031	0.037
Race	-0.025	-0.024	-0.056	-0.023	0.032	0.037
6. Alternative value-added models						
Homogeneous	0.006	0.009	-0.018	0.003	0.056	0.067
Using school means	0.129	0.121	0.099	0.142	0.202	0.206
Using alternative FEs	-0.006	-0.003	-0.029	-0.003	0.047	0.052
7. Allow for correlated random coefficients in teacher preferences						
Corr. R.C.	-0.020	-0.019	-0.052	-0.020	0.033	0.037
8. Vary teacher preference specification to use binary logit						
No REs or FEs	-0.018	-0.017	-0.040	-0.020	0.033	0.037
Teacher REs, School FEs	-0.021	-0.022	-0.045	-0.020	0.033	0.037
Teacher FEs, School FEs	-0.022	-0.022	-0.046	-0.020	0.033	0.037
9. Add covariates to principal model						
	-0.0177	-0.0150	-0.047	-0.020	0.037	0.037
10. Multiply value-added coefficients by 10 in principal model						
	-0.015	-0.014	-0.047	-0.025	0.033	0.037

The table shows robustness checks for our main results. The columns show the value-added of teachers assigned to disadvantaged students. The allocations correspond to those described in Figure 4, and the variants are described in the text.

istic, z_{jpt}. Teacher j's utility for teaching at position p in year t is

$$\tilde{u}_{jpt} = u_{jpt} + \gamma(b_0 + b_1 z_{jpt}), \tag{15}$$

where we multiply by the commute time coefficient (γ) to express bonus spending in minutes of commute time. 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.[31]

To focus on policies that are likely to receive teachers' support, we make each teacher weakly better off than in the status quo equilibrium. We set the transfer such that the teacher with the worst change is indifferent. This lump-sum transfer can be either positive or negative. Thus, the district's total bonus to a teacher depends both on the choice of how much to compensate for the characteristic and how it changes the allocation.[32]

We examine bonus schemes over two objects. First, we study bonuses based on the number of disadvantaged students the teacher has ($n_{k(p)1t}$). These bonuses mimic the hard-to-staff school bonuses that some districts have piloted. Second, we interact school and teacher characteristics by considering bonuses based on a teacher's absolute advantage times the number of disadvantaged students: $((p_{0t}\hat{\mu}_{j0t} + (1 - p_{0t})\hat{\mu}_{j1t})n_{k(p)1t})$.

8.2 Results

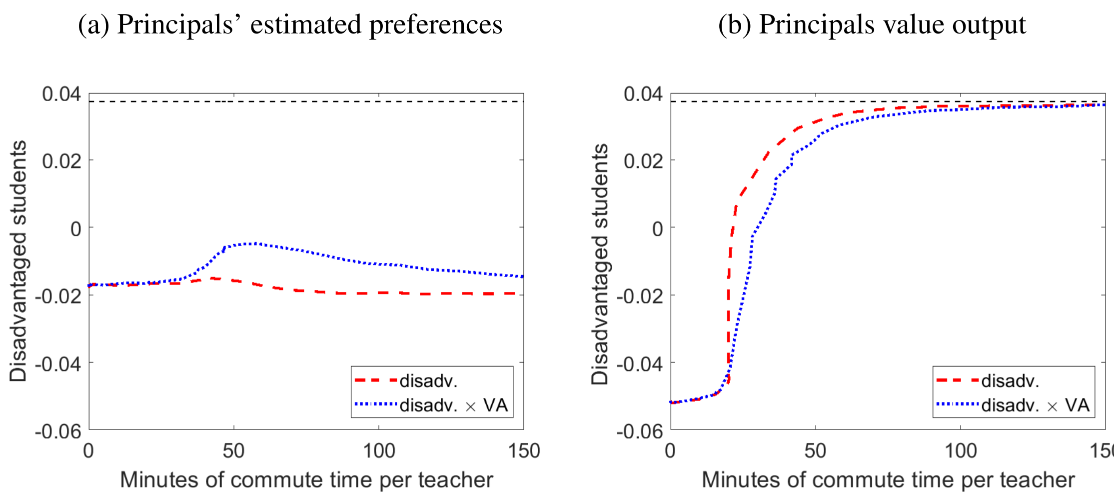
Panel (a) of Figure 5 shows the effect of these two bonus schemes on disadvantaged students' test scores when principals hire according to their estimated preferences. The top line shows achievement in the the first-best allocation. To allow for comparisons across bonus schemes, the horizontal axis is the total realized spending (normalized to be in minutes of commute time per teacher).

We have three results, all of which reflect the theory of the second best. First, untargeted

[31]We compare the effectiveness of bonuses with equivalent utility costs. Because we use the same conversion factor for all schemes, the conversion factor does not affect the comparisons.

[32]Formally, let $\Delta\tilde{u}^{b_1}_{jpt} = (u_{jp^*(j)t} - u_{jpt}) + \gamma b_1 z_{jp^*(j)t}$ be the change in teacher j's utility (excluding the transfer) between the zero-bonus and the b_1 bonus equilibria. The transfer is: $b_0 = -\min_j \Delta\tilde{u}^{b_1}_{jpt}$. The total bonus to teacher j is $b_0 + b_1 z_{jp^*(j)t}$.

Figure 5: Teacher bonus schemes



This figure shows the effect of teacher bonus schemes on the achievement of disadvantaged students. 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 disadvantaged student. We consider two policies: subsidizing the position based on the number of disadvantaged students in the position, and subsidizing the position based on number disadvantaged interacted with the teacher's absolute advantage. In the left panel, we take as the baseline allocation the status quo. In the right panel, we take as the baseline allocation one where principals maximize output. The horizontal dashed lines show the output in the first-best.

bonuses for teaching disadvantaged students are ineffective, and even slightly harmful, in raising disadvantaged students' test scores. As in the last section, shrinking one deviation need not move the allocation closer to the first-best. Second, targeted bonuses that pay the best teachers more for teaching disadvantaged students are more effective than untargeted bonuses because they jointly address deviations on both sides of the market. On the one hand, targeting the bonus does less to align preferences with the planner because only a subset of teachers change their preference. On the other hand, the targeting substitutes for the deviation on the other side of the market: if the applicant pool is steered toward the best teachers, then principals' deviations matter less. Third, the bonuses eventually become less effective as they grow larger. Here, larger bonuses expand the applicant pool for disadvantaged schools, but the larger pool causes the deviation in principal preferences to matter more.

Effective policy needs to address the deviations jointly. In Panel (b) of Figure 5 we consider the effect of the teacher bonus schemes when principals hire according to value-added. Such hiring rules may be induced by a combination of an information intervention and principal bonuses for hiring effective teachers.

We again have three results. First, as in the prior section, if teacher bonuses are small such that estimated teacher preferences largely guide applications, then principals hiring according to value-added leads to large decreases in disadvantaged students' test scores. Fixing the deviation on the principal side, but hardly closing the teacher deviation, has a large negative effect. Second, as teacher bonuses get larger, principals hiring according to value-added make the teacher bonuses particularly effective. At the equivalent of about 50 minutes of commute time per teacher, the bonuses have nearly reached the first-best. That teacher bonus effectiveness is increasing in principal bonuses (or information interventions) reflects the interaction of the two sides of the market. We estimate that once average teacher bonuses exceed about 25 minutes of commute time, disadvantaged students do better when principals hire according to value-added. Third, untargeted bonuses now outperform targeted bonuses at nearly all cost levels. Because the principal deviation has been closed, the targeting of bonuses is no longer needed. In fact, such targeting is now counter-productive.

9 Discussion

We have studied the equity consequences of the within-district allocation of teachers to schools. We consider both the current allocation and alternative policies. To decentralize the first-best that maximizes disadvantaged students' achievement, teachers would need to prefer schools with more disadvantaged students and principals would need to prefer higher value-added teachers. Using rich data from the teacher transfer system that allows us to observe actions, we show that both sides' preferences deviate from these. Nonetheless, and consistent with the theory of the second best, these two deviations interact to generate a surprisingly equitable allocation, where disadvantaged students do not have worse teachers than advantaged students. In terms of policy, and again consistent with the theory of the second best, fixing one deviation at a time is either ineffective or harmful. Fixing both deviations could close about a fifteenth of the achievement gap per year.

More broadly, this paper has demonstrated the value of using rich data to study the functioning of particular labor markets. 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 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

Agarwal, Nikhil. 2017. "Policy analysis in matching markets." *American Economic Review, Papers and Proceedings* 107 (5):246–50.

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

Angrist, Joshua, Peter Hull, Parag A. Pathak, and Christopher Walters. Forthcoming. "Race and the Mismeasure of School Quality." *American Economic Review: Insights* .

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.

Aucejo, Esteban M, Patrick Coate, Jane Cooley Fruehwirth, Sean Kelly, and Zachary Mozenter. 2022. "Teacher Effectiveness and Classroom Composition: Understanding Match Effects in the Classroom." *Economic Journal* 132 (648):3047–3064.

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

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.

Bates, Michael D, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin. 2022. "Teacher Labor Market Equilibrium and Student Achievement." Working Paper 29728, National Bureau of Economic Research.

Bau, Natalie. 2022. "Estimating an equilibrium model of horizontal competition in education." *Journal of Political Economy* 130 (7):1717–1764.

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

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 29068, National Bureau of Economic Research.

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, 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.

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.

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.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *The Quarterly Journal of Economics* 128 (3):967–1015.

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. 2014a. "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. 2014b. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review* 104 (9):2633–79.

Chiappori, Pierre-André. 2020. "The theory and empirics of the marriage market." *Annual Review of Economics* 12:547–578.

Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor. 2011. "Teacher mobility, school segregation, and pay-based policies to level the playing field." *Education Finance and Policy* 6 (3):399–438.

Combe, Julien, Umut Dur, Olivier Tercieux, Camille Terrier, and M Utku Unver. 2022. "Market Design for Distributional Objectives in (Re) assignment: An Application to Improve the Distribution of Teachers in Schools." Working paper.

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

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." Working paper.

Dahlstrand, Amanda. 2022. "Defying Distance? The Provision of Services in the Digital Age." Working paper.

Davis, Jonathan. 2022. "The Short and Long Run Impacts of the Deferred Acceptance Algorithm: Evidence from Matching Teach For America Teachers to Schools." Working paper.

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

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

Delgado, William. 2022. "Heterogeneous Teacher Effects, Comparative Advantage, and Match Quality." Working paper.

Diamond, William and Nikhil Agarwal. 2017. "Latent indices in assortative matching models." *Quantitative Economics* 8 (3):685–728.

Ederer, Tim. 2022. "Labor market dynamics and teacher spatial sorting." Tech. rep.

Elacqua, Gregory, Anne Sofie Westh Olsen, Santiago Velez-Ferro et al. 2021. "The Market Design Approach to Teacher Assignment: Evidence from Ecuador." Tech. rep., Inter-American Development Bank.

Engel, Mimi, Marisa Cannata, and F Chris Curran. 2017. "Principal influence in teacher hiring: Documenting decentralization over time." *Journal of Educational Administration* 56 (3):277–296.

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.

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, Vanessa Quince, and Roddy Theobald. 2018. "Has it always been this way? Tracing the evolution of teacher quality gaps in US public schools." *American Educational Research Journal* 55 (1):171–201.

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.

Greenberg, David and John McCall. 1974. "Teacher mobility and allocation." *Journal of Human Resources* 9 (4):480–502.

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.

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.

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

———. 2018b. "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." Working Paper 14607, National Bureau of Economic Research.

Kapor, Adam. 2020. "Distributional effects of race-blind affirmative action." Working paper.

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.

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.

Mundlak, Yair. 1978. “On the pooling of time series and cross section data.” *Econometrica* 46 (1):69–85.

National Council on Teacher Quality. 2022. “Teacher contract database.”

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.

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

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

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.

Silhol, Julien and Lionel Wilner. 2022. “Teachers’ Desired Mobility to Disadvantaged Schools: Do Financial Incentives Matter?” Working Paper 2022-9906, CESifo.

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

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

Online Appendix

Teacher labor market policy and the theory of the second best

Michael Bates, Michael Dinerstein, Andrew Johnston, and Isaac Sorkin

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 the 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 in 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 the last four digits of social security numbers, names, and birth dates. For teachers who had a sufficiently good match, 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 vary 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 restriction 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 Do teachers bargain over student assignment?

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.

In Appendix Table A11, there is a strong relationship between experience and student attributes. More experienced teachers are assigned higher-scoring students, fewer economically disdavantaged students, more students designated as gifted, and fewer Black students.

Within-school, however, more experienced new teachers are not assigned more "preferable" classroom compositions in that the relationships are an order of magnitude weaker than the unconditional relationship, and sometimes statistically insignificant.

C Selection into the transfer market

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 A10 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.

D Omitted details on value-added model

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

Here we formally state the assumptions that were informally discussed in Section 4.

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:*

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

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

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:

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

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

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:*

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

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

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

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

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:*

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

D.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, $\hat{\nu}_{it}$:[33]

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

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:

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

[33] 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.

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 (2022) 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 (2022). Here we describe how we estimate the structural parameters: $\sigma_{\varepsilon 0}, \sigma_{\varepsilon 1}, \sigma_{\theta 0}, \sigma_{\theta 1}$, $cov(\theta_{c0t}, \theta_{c1t}), \sigma_{\mu 0}, \sigma_{\mu 1}$, $cov(\mu_{j0t}, \mu_{j1t}), cov(\mu_{j0t}, \mu_{j0s})$, $cov(\mu_{j1t}, \mu_{j1s}), cov(\mu_{j0t}, \mu_{j1s})$.

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

- $\hat{cov}(\mu_{j1t}, \mu_{j1s}) = cov(\bar{A}_{j1t}, \bar{A}_{j1s})$
- $\hat{cov}(\mu_{j0t}, \mu_{j1s}) = cov(\bar{A}_{j0t}, \bar{A}_{j1s})$

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.[34] 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. In our setting many elementary school teachers have students from multiple classes. The prevalence of multiple classrooms is increasing over time (Appendix Table A13). 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.

Appendix Table A14 presents parameter estimates. 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

[34]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.

with the two types of students. We find large returns to experience in the first year, and then a profile that flattens out after about four years of experience.

D.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}).

D.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. Beyond the exercise presented in Appendix Figure A2, we present two additional ways of testing 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 95% confidence interval is $(0.73, 0.87)$ (Appendix Table A14).

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 ($p < 0.0001$).[35]

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. If our estimated comparative advantage is meaningful, 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. Table A12 shows that for teachers with a comparative advantage in teaching advantaged students, 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 indicate that comparative advantage is persistent across settings and predictive of match-specific productivity.

D.5 Behavioral value-added

In robustness checks, we incorporate a measure of a teacher's value-added on behavioral outcomes (Jackson, 2018). Because we focus on elementary school teachers, we have fewer outcomes available (e.g., no grades). We thus measure teachers' effects on a student's log absence rate, whether the student has any in-school suspension, and whether the student has any out-of-school suspension. We recover the first principal component and use this as

[35] 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.

our outcome.

We estimate two-dimensional behavioral value-added with identical methods to those we use for math value-added. When controlling for lagged student outcomes, we use the lagged value of the first principal component.

E 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:

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

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: 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 A3: 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 A4: 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 A5: 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 A6: 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 A7: 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 A8: Summary statistics for 2015-16, by economic disadvantage

	Focal, Adv	Focal, Disadv	Other, Adv	Other, Disadv
Students				
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
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 or more years	38.44	29.47	51.83	46.33
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
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
Mean behavioral value-added				
Baseline	0.02	0.02		

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 A9: Potential Gains from Reassignment

	Non-Disadvantaged	Disadvantaged
Targeting Disadvantaged Students		
Max Disadvantaged VA	-0.049	0.075
Max Disadvantaged VA, Robustness		
Constant Class Size	-0.111	0.096
Homogeneous VA	-0.081	0.076
Using School Means in VA	-0.056	0.084
Using Alternative FEs in VA	-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 matching better teachers to disadvantaged students. Gains are measured in student standard deviations (σ). The first and second columns show the per-student gains, relative to the actual allocation, for non-disadvantaged and disadvantaged students. '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 "Constant Class Size" allocation.

Table A10: 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.

Table A11: 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

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 A12: 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 comparative advantage in pre-transfer schools. 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 A13: 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 A14: Teacher Value-Added Structural Parameters

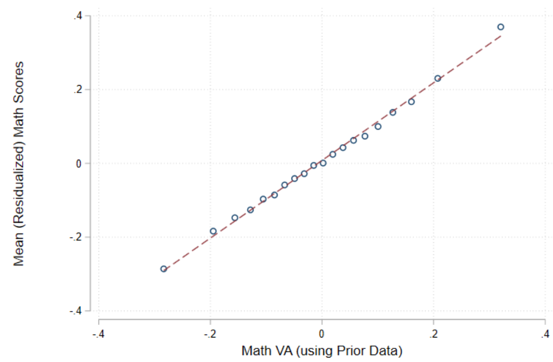
	Estimates	Standard Errors	95% CI Lower Bound	95% CI Upper Bound
$\sigma_{\varepsilon 0}$	0.450	0.000	0.456	0.457
$\sigma_{\varepsilon 1}$	0.470	0.000	0.477	0.479
$\sigma_{\theta 0}$	0.110	0.007	0.108	0.137
$\sigma_{\theta 1}$	0.088	0.015	0.089	0.143
correlation$(\theta_{c0t}, \theta_{c1t})$	0.657	0.162	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

	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

	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

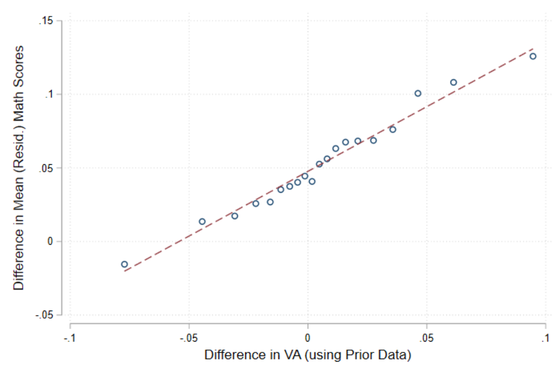
In Panel A, the table shows the estimates of a subset of the structural parameters of the production model – specifically the parameters corresponding to contemporaneous output. Non-disadvantaged students have index 1 while disadvantaged students have index 2. ε is the student-year idiosyncratic component, 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. In Panel B, 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. 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.

Figure A1: Math Value-Added Forecast Unbiasedness



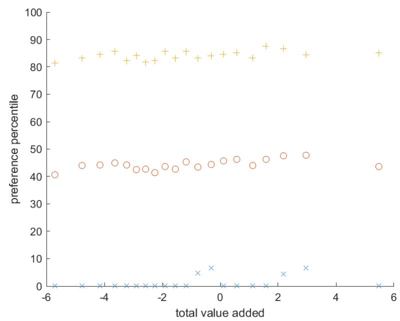
The figure is a binscatter, where an observation is a teacher-year and math value-added estimates are predictions using data from prior years. Units are student standard deviations. The y-axis is the mean student math test score, residualized by student demographics including lagged scores, school fixed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

Figure A2: 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 A3: 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.