NBER WORKING PAPER SERIES

PREFERENCES, SELECTION, AND THE STRUCTURE OF TEACHER PAY

Andrew C. Johnston

Working Paper 33088
http://www.nber.org/papers/w33088

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2024

I am grateful to the Editor, Christopher Walters, and to three anonymous referees for excellent comments and suggestions. The paper has been enriched through conversation with Michael Bates, Barbara Biasi, David Card, Damon Clark, Gordon Dahl, Michael Dinerstein, Mark Duggan, Laura Giuliano, Eric Hanushek, Caroline Hoxby, Alex Mas, Nolan Pope, Jesse Rothstein, Kenneth Shores, and Isaac Sorkin, as well as seminar participants at the National Bureau of Economic Research, Princeton University, UC Berkeley, and Stanford University. Benjamin Feis provided outstanding research assistance. Financial support from the Institute for Education Science, the U.S. Department of Labor, and National Bureau of Economic Research is gratefully acknowledged. All errors my own. Johnston, 2024. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Preferences, Selection, and the Structure of Teacher Pay
Andrew C. Johnston
NBER Working Paper No. 33088
October 2024
JEL No. I20, J32, J45, M50

ABSTRACT

I examine teacher preferences using a discrete-choice experiment, which I link to administrative data on teacher effectiveness. I estimate willingness-to-pay for a rich set of compensation elements and working conditions. Highly effective teachers usually have the same preferences as their peers, but they have stronger preferences for performance pay. I use the preference estimates to investigate the optimal compensation structure for three key objectives: maximizing teacher utility, maximizing teacher retention, and maximizing student achievement. Under each objective, schools underutilize salary and performance pay, while overutilizing retirement benefits. Restructuring compensation can significantly improve both teacher welfare and student achievement.

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

A data appendix is available at http://www.nber.org/data-appendix/w33088

I. Introduction

Human capital is a fundamental driver of income, inequality, and economic growth (Neal and Johnson 1996; Barro 2001; Chetty, Friedman, and Rockoff 2014b). In its formation, teacher quality emerges as arguably the most important public input (Darling-Hammond 2003; Rockoff 2004; Rivkin, Hanushek, and Kain 2005). Highly effective teachers not only boost achievement but also develop essential non-cognitive skills, resulting in higher lifetime earnings and improved adult outcomes for their students (Chetty, Friedman, and Rockoff 2014; Opper 2019; Gilraine and Pope 2020). Teachers vary considerably in their effectiveness, and they improve with experience. These facts imply the focal importance of teacher *selection* and *retention* in building the stock of human capital.

In the United States, governments spend some $1 trillion per year on K–12 education, the principal cost of which is *compensation* (80 percent). Teachers receive a distinctive pay structure: low salary, generous retirements, and few performance incentives. The effectiveness of this structure is questionable, for two reasons. First, public schools possess significant market power and no market test, so schools have little incentive to structure compensation optimally (Hoxby 2000; Rothstein 2007; de Ree et al. 2018). Second, political incentives may further distort compensation away from what is optimal (Hoxby 1996; Clemens and Cutler 2014; Glaeser and Ponzetto 2014; Fitzpatrick 2015; Lovenheim and Willen 2019). The question I address in this paper is whether substantial investments in teacher compensation are structured *well.*

I shed light on the question in three steps. First, I estimate teacher preferences for compensation structure and working conditions using a discrete-choice experiment. Schools can improve the appeal of teaching by shifting resources toward features teachers value most relative to cost. Second, I investigate whether high-performing teachers exhibit distinct preferences for pay structure and working conditions. If so, policymakers can improve selection by tailoring compensation to appeal to high performers. Third, I leverage these preference estimates to model counterfactual policies that maximize key policy objectives within the current budget constraint. The model allows me to identify opportunities for efficiency gains.

Measuring preferences in the teacher labor market presents unique challenges. While economists typically infer preferences from observed choices among available options (Train

2009; Wiswall and Zafar 2017), this approach is difficult when studying teachers. First, teachers rarely if ever entertain multiple job offers at once.[2] Second, observable characteristics of schools are probably correlated with important unobservable ones (Antos and Rosen, 1975). And, perhaps most critically, third: public schools exhibit minimal variation in compensation and working conditions. Employment contracts are essentially uniform across schools, and many important features of the policy space do not vary at all.[3]

To overcome these challenges, I conducted a discrete-choice experiment in Aldine Independent School District, a large urban district near Houston. The experiment presented K–12 teachers with a series of hypothetical job offers to choose between. While hypothetical choices often provide weak incentives for truthful responses, teachers had reason to reveal their true preferences in this setting. The district framed the survey as a tool to inform new personnel policies, effectively positioning each question as a referendum between compensation proposals (Carson et al. 2000). This approach, combined with a high response rate of 97 percent, lends credibility to the results. The job offers in the experiment varied multiple dimensions of compensation and working conditions, including: starting salary, salary growth, retirement generosity, defined-benefit or defined-contribution plans, health insurance premia and deductibles, performance pay, the basis of evaluation, time-to-tenure, class sizes, administrative support, student demographics, and principal support.

The discrete-choice experiment allows me to estimate the structure of teacher preferences for multiple job attributes simultaneously. As a result, I can model the optimal compensation structure—something one cannot do one by one from isolated natural experiments.

The responses from the experiment align with several available benchmarks for realism, lending credibility to the findings (Mas and Pallais 2017; Wiswall and Zafar 2017; Maestas et al. 2018; Hainmueller et al. 2014). For instance, if teachers pay part of their health-insurance premium, they should be indifferent between an additional dollar of salary and an additional dollar offsetting what they pay for insurance. Reassuringly, teachers value health-insurance subsidies *identically* to an equivalent increase in salary.[4] Further, the discount rate implied by

[2] Teachers rarely entertain simultaneous offers since they are spread out over eight months of each year and usually explode on the same day they are extended.
[3] State policy and common union influence generate similar compensation structures across districts. Within district, compensation is totally uniform.
[4] This is especially remarkable because the two quantities are presented in different units, yearly salary versus monthly premia.

teacher choices between salary and retirement benefits aligns with estimates from the literature on time preferences (Best et al. 2018; Ericson and Laibson 2018). Finally, the implied cost of commuting derived from teacher choices matches estimates from an established urban literature (Small 2012; Mas and Pallais 2017). These validations against known benchmarks instill confidence in the more novel estimates.[5]

To understand how teachers value different components of their workplace, I use the choice experiment to calculate willingness-to-pay (WTP) for each of a rich set of attributes. Teachers have concave preferences for performance pay and convex disutility from class size. A ten-student reduction in class size (from 30 to 20 students) is valued equivalent to a $6,000 increase in salary.[6] When retirement income is held constant, teachers prefer portable, defined contributions plans over traditional pensions. And an additional year of probationary status before tenure is equivalent to a salary reduction of 1 percent.

These findings are new to the literature, and I provide novel estimates on WTP for a broad array of other attributes. Among those considered, teachers most value a principal who is "supportive" in handling disruptive students, as opposed to "hands-off." They value this attribute as much as a 17-percent salary increase. A supportive principal also significantly mitigates teacher aversion to challenging work environments. It eliminates 90 percent of teachers' disutility to work in low-achieving schools and 85 percent of the disutility of teaching in low-income schools. This suggests that attentive managers reduce the perceived cost of teaching in hard-to-staff schools.

Policymakers can enhance the appeal of teaching by shifting resources into compensation elements and work amenities that teachers value highly relative to cost. I find that teachers value defined-benefit retirement income at 60 percent of its cost. In contrast, they value defined-contributions retirement income at 104 percent of its cost.[7] The cost of reducing class size is seven times greater than teachers' willingness-to-pay (WTP). Teacher's WTP for health-insurance subsidies is equal to the cost of provision. And, on average, teachers value a

---

[5] I also test whether utility estimates reflect real-world preferences. Group preference for the status quo bundle (e.g., male, female, black, white, Hispanic, and quartiles of tenure) predicts group attrition rates with a correlation of 0.8. This further reinforces the claim that the preferences teachers disclose are those animating relevant labor-supply decisions.
[6] Here, base starting pay is $50,000 for a new teacher with a bachelor's degree in 2016 dollars.
[7] An additional ten-point replacement rate in pension is valued equivalent to a $1,730 salary increase, less than its cost of $2,870 per year. The cost of the same expected retirement income through a defined-contributions plan is $1,663.

unit of performance pay a third more than its cost. These findings suggest opportunities to improve teacher satisfaction and retention without increasing overall cost.

Predicting teacher effectiveness from job applications remains a significant challenge (Rockoff et al., 2011; Jacob et al., 2018).[8] However, if high-performing teachers have distinct preferences, policymakers can influence selection by structuring compensation to appeal to high performers (Ballou 1996; Hanushek 2011).[9] To investigate this, I estimate each teacher's contribution to student learning using value-add (VA) scores derived from student achievement records. High-performing teachers generally share similar preferences with their colleagues, with one notable exception. Excellent teachers prefer jobs that include performance pay. A top-decile teacher is willing to pay $510 for a performance pay contract of $1,000 to the top quarter of teachers, while a bottom-decile teacher is willing to pay $83. As a result, top-decile teachers are 23 percent more likely than bottom-decile ones to choose an offer providing $3,000 in performance pay. This preference difference implies performance pay induces favorable selection.[10] Importantly, the district does not estimate or disseminate VA scores to teachers, indicating that teachers likely have private information about their own productivity.

Using the preference estimates, I investigate the optimal pay structure for teachers. Specifically, how should schools design compensation to achieve three key objectives: (1) maximizing teacher welfare, (2) maximizing teacher retention, and (3) maximizing student achievement. To conduct this analysis, I map teacher utility to retention decisions and calibrate a model of the achievement production function.[11] A consistent pattern emerges. Whether maximizing teacher utility, teacher retention, or student achievement, schools appear to underutilize salary and performance pay, while overutilizing retirement benefits. This suggests the current structure is suboptimal for achieving any of these core goals.

[8] See also Hanushek (1986), (1997), Greenwald et al. (1996), and Sajjadiani et al. 2019.

[9] Over time, the effect may be especially pronounced since the preferred compensation differentially retains high-performing teachers who may themselves be a draw and productive input to other teachers (Feng and Sass 2017; Papay et al. 2020). Raising everyone's compensation may improve the average quality of new recruits, but it reduces the scope for new hiring since ineffective teachers are also more likely to be retained (Ballou 1996).

[10] It is less clear whether merit pay would affect sorting *into* the profession if prospective teachers cannot forecast their teaching productivity.

[11] It is useful to notice that the problems nest one another. Achievement is a function of teacher retention, and teacher retention is a function of teacher welfare.

When optimizing teacher compensation to improve teacher welfare, the optimal structure increases teacher welfare by 20 percent—equivalent to giving each teacher a permanent $17,000 raise. Doing so increases teacher retention substantially, and student learning rises by 0.04 student standard deviations (σ) per year. Maximizing teacher retention prescribes a similar structure but increases salary growth to improve retention for experienced teachers. This retention-focused structure raises the average teacher experience by 22 percent in equilibrium and increases the odds of a student having a veteran teacher by a third.[12] These findings demonstrate that appealing to teacher preferences can also improve student outcomes.[13]

The compensation structure that maximizes achievement shares similarities with the retention-maximizing structure, but incorporates more performance pay. The achievement-optimal structure leads to a substantial improvement in student learning, increasing achievement by 0.09 standard deviations per year.[14] Achievement gains are primarily driven by performance pay, improving selection and effort, with improved overall retention accounting for 10 percent of the increase. This structure also offers a Pareto improvement across agent types: it not only increases student achievement, but also improves teacher welfare by 7 percent over the *status quo*.

The baseline simulation calculates the partial equilibrium effect, estimating what would happen if a single district implemented the optimal reform. To assess the general equilibrium effect—the system-wide impact if all school districts implemented it—I leverage changes in salary schedules as a natural experiment (Hendricks, 2014). This approach allows me to distinguish between retention gains from prevented transfers to other districts, and those from prevented professional retirements. The latter benefits the entire school system, while the former does not. It only redistributes talent across schools.

The results of the natural experiment indicate that 87 percent of retention gains from salary increases are attributed to professional retentions (highly significant), while only 13

[12] The similarity between utility-maximizing and retention-maximizing structures underscores the close relationship between teacher welfare and longevity in the profession.
[13] I classify teachers in the top quarter of the experience distribution as "veterans" which is having more than 15 years of experience.
[14] For comparison, these are about two-thirds the size of the effect of increasing teacher value-add by 1 standard deviation (Chetty, Friedman, and Rockoff, 2014) or about a third larger than the effect of adopting the most effective text book (Koedel et al., 2017).

percent come at the expense of other districts (statistically insignificant). This finding has two important implications. First, it suggests substantial local market power by school districts. Second, it indicates that the general equilibrium effect would be nearly identical to the partial equilibrium one—differing by just 1.6 percent.

The counterfactual analysis assumes that marginal teachers have similar preferences to inframarginal ones.[15] To assess this assumption, I conducted two direct examinations of marginal teacher preferences. First, I compared the preferences of teachers who eventually leave the district with those who stay. Second, I surveyed college students and analyze whether preferences differ based on their interest in teaching. In both cases, the preferences of marginal teachers are indistinguishable from inframarginal ones, suggesting that the counterfactuals are robust.[16] Importantly, this finding also implies that teachers have quite similar preferences for compensation structure as other workers, they only differ in their taste for *teaching*.

This study contributes to several literatures exploring teacher preferences, compensation, and selection. For preferences, Boyd et al. (2013) estimate teacher preferences for school characteristics from equilibrium matches, and Bates et al. (2022) derive preference estimates from detailed teacher application choices. Regarding compensation, Hendricks (2014) exploits changes in salary schedules to estimate the retentive effects of compensation. Fitzpatrick (2015) and Biasi (2019) both exploit reforms to measure how teachers value their pensions.

The impact of performance incentives on teacher quality and selection has been examined in various contexts. Rothstein (2015) simulates the effect of performance incentives on labor supply. Biasi (2021) demonstrates that allowing principals to distribute performance pay leads to improved selection and effort among teachers, with Burgess et al. (2022) reporting similar findings.[17] In the developing world, Leaver et al. (2021) and Brown and Andrabi (2021) find that performance pay improves both effort and selection.

[15] By "marginal teacher" I mean a teacher or college student that is nearly indifferent to teaching. Without similar preferences, the simulations may be inaccurate for two reasons. First, the solution may not be stable if a counterfactual policy changes the composition of teacher preferences. Second, retention will be poorly modeled if potential leavers have dissimilar preferences to inframarginal teachers.

[16] Specifically, the results are robust to changing selection on entry and exit since preferences are constant along the interest-in-teaching dimension. The similar preferences of leavers to stayers is also helpful because the model calculates exit rates based on average preferences. If teachers on the leaving margin had distinct preferences, my retention and cost calculations would be wrong.

[17] Willen (2021) also examines the effect of a reform that allowed principals to individuate teacher pay, and he finds quite different results. Whereas principals in Wisconsin used their new-found flexibility to reward excellent instructors, principals in Sweden did not.

This study makes three key contributions to the existing literature. First, it provides comprehensive estimates of willingness-to-pay (WTP) for a wide array of compensation vehicles and working conditions using a choice experiment. This approach allows for the evaluation of factors that don't typically vary independently in real-world settings, addressing a significant limitation in prior research. Second, it directly examines the relationship between worker preferences and productivity, offering insights into how compensation structures might affect selection. Third, it uses these preference estimates to conduct counterfactual policy experiments grounded in real data, providing a more robust basis for policy recommendations.

II. Experimental Design and Econometric Framework

Implementation

I construct a survey that invites teachers to make a series of choices between hypothetical job offers. To increase power, I use the statistical package, JMP, that varies the attributes using a fractional conjoint design assuming additive separability. Each choice set requires the teacher to make tradeoffs, and the package maximizes efficiency of the parameters of the utility model for a given number of choice sets.[18] The choice experiment allows the analyst to evaluate several hypotheses in a single study and compare the influence of various factors within a shared setting, making the estimates directly comparable and useful for understanding tradeoffs in counterfactuals.

In 2016, Aldine Independent School District (ISD) was interested in revamping their personnel policies to improve retention. Aldine included my questions in their yearly survey to teachers to understand: *How should they structure personnel policy?* I refer to the district goal in the survey invitation and instructions for respondents. The invitation reads: "Please take a few minutes to respond to Aldine's annual survey. Your insights will be helpful as we improve policies to meet your needs." The instructions wrote, "The purpose of the survey is to learn more about the attitudes of teachers…to improve Aldine's teacher policies." Because my questions were embedded in the district's yearly survey to teachers, teachers could infer the district was interested in knowing their preferences over decisions ranging from class size

[18] I assume, for instance, that teachers prefer more of each type of compensation (higher staring salary, greater salary growth, a more generous retirement, etc.) while assuming that teachers prefer less of other things (e.g., fewer students to a class, shorter probationary period, smaller student-poverty shares, etc.). The software generates choice sets that present tradeoffs between attributes that are assumed to be desirable. The compensation questions present options that are about equally costly.

to salary. Responses were anonymized to the district. After the survey, I presented the results to the district.[19]

Survey questions feature fourteen attributes. These include (1) starting salary, (2) salary growth rate, (3) health insurance plan (in terms of the deductible and monthly premium), (4) retirement benefits (in terms of the expected replacement rate and the mode, either a defined benefits (DB) or defined contribution (DC) plan), (5) performance-pay program, (6) class size, (7) the duration of the probationary contract (essentially "time-to-tenure"), (8) the frequency of contract review and renewal, (9) how many hours of teaching assistance a school provides the teacher, (10) the percent of students who are low income, (11) the percent of students who are minorities, (12) the average achievement percentile of students, (13) commuting distance in time, and (14) whether the principal is "supportive" or "hands-off" with disruptive students. Attributes take on several values, shown in online Appendix table 1.[20]

When constructing the survey, the analyst faces a tradeoff between the realism of the options (made richer in the number and detail of attributes) and the ability of respondents to compute their preferences in a short time. If the attributes are too numerous (generally considered more than 10 in a single choice (Hainmueller et al., 2014)), respondents tend to resort to a simplifying rule in which they consider a subset of attributes they find most important. To estimate preferences over many factors, I split the attributes into three sets of questions, called "decks."

The first deck asks teachers to choose between different compensation structures, varying starting salary, salary growth rate, health insurance subsidies, retirement plans, and merit compensation. I include the starting-salary attribute in each of the other decks to "bridge" the decks, allowing for preference comparisons between attributes in different decks. The second deck varies working conditions, including class size, how long new teachers are on

[19] The district indicated that they incorporated the research as they set new policy, but the changes were not dramatic. They increased salaries by 4.5 percent over the next two years, somewhat above inflation over the same period (3.4 percent). They also attempted to implement a performance pay regime in response to the study, but the initiative stalled when they could not find a "fair" way to implement performance pay, since some teachers were in tested subjects and others were not.

[20] Some of these features change in more than one dimension. For instance, the retirement description varies the replacement rate the plan provides in expectation and whether retirement is based on a defined-contribution or a traditional, defined-benefit plan (essentially the difference between a 401(k) and a pension). The health insurance description varied how much the district paid, the deductible, and the copay for an office visit. The performance-pay attribute varied how much a teacher could receive for being in the top 25 percent of teachers, either based on student growth and principal evaluations or student growth alone.

probationary contracts, how often term contracts are reviewed and renewed, distance to work from home in travel time, and how many hours of instructional support are provided the teacher each week. The third asks teachers to choose between job offers that vary starting salary (again, to assimilate estimates across decks), student poverty share, student minority share, average achievement percentile, and whether a principal was “supportive” or “hands-off” with disruptive students, as well as a placebo attribute. The statistical software generated 30 questions for each of the three decks and respondents were presented, at random, four questions from the compensation deck, four questions from the working-conditions deck, and three questions from the student and principal characteristic deck, since the final deck had fewer parameters to estimate. Examples of these survey questions are presented in online Appendix figures 1–3.

A variety of features of the experiment conduce truthful responses in hypothetical choice. (1) Hypothetical choices with *tradeoffs* align with real-world choices. (2) Recent studies in labor find the career preferences elicited in hypothetical settings match those elicited in incentivized ones. (3) Hypothetical choices where people have experience with the context produce truthful responses. (4) The actual preferences elicited in my experiment closely match the theoretical and empirical benchmarks available. (5) Social-desirability bias is credibly avoided since choices are private and many attributes vary at once. And (6), in this setting, there is *consequence* to teachers’ choices because each question represents a referendum, where the district gathers information about how to adjust personnel policy. I expand on these points and provide support from the literature in Appendix A for interested readers.

An important criticism of conjoint experiments is that, by asking subjects to make tradeoffs between options, the researcher implicitly designates as valuable attributes subjects might not care about in real life—a type of Hawthorne effect that results in upward-biased estimates of unimportant items. To examine this concern, I include a placebo feature that should have no plausible impact on teacher utility—whether the school bus at the featured school is blue (McFadden 1981)—to evaluate whether the experimental setting encourages teachers to exhibit preferences for things that have no impact on their welfare. Teachers express no preference over this irrelevant detail, aiding interpretation. Uninstructed, subjects may fill in the state space, inferring characteristics that influence their choice other than the features explicitly described. I frame each question by asking teachers to imagine that two hypothetical job offers are identical in every other way, following Wiswall and Zafar (2017): "*If two schools that were identical in every other way made the following offers, which would you prefer*?"

Inattention is not a major issue. First, inattention that is not correlated with the attributes generates random attribution error in the outcome variable, and does not bias estimates (Wooldridge 2010).[21] Second, the survey is administered digitally, and the option to advance to the next question does not appear for the first five seconds each question is available, nudging teachers to read the prompt as they wait for an unstated amount of time. Third, the online survey environment records how long each respondent takes to respond to each question; teachers take on average 35 seconds per question. I estimate the models separately among respondents who took longer-than-average and shorter-than-average times to respond. Preference estimates are essentially identical in the two subsamples, suggesting time-to-respond does not affect the estimates.[22]

I deployed the experiment in a large, urban school district called Aldine Independent School District in Texas, at end of the school year in May 2016. I invited each of the district's 4,358 teachers to participate in the experiment, 97.8 percent of whom completed the survey. The high response rate was encouraged by district support (the survey was part of their yearly feedback from teachers), reminder emails each day during the week of the survey, and a lottery for $10 gift cards for those that completed the survey.[23]

Conceptual and Econometric Framework

I use teacher's hypothetical choices to estimate canonical utility models (Louviere 2000; Train 2009; Zafar and Wiswall 2017). Each selection requires teachers to make tradeoffs between features that are assumed to generate positive utility. Under weak conditions, the hypothetical job selection data identify job preferences over several factors while standard realized match data do not. Teacher i chooses offer a if $U_i(\vec{c}_a, \vec{w}_a) > U_i(\vec{c}_b, \vec{w}_b)$, where $\vec{c}_x$ represents a vector describing the compensation structure of option $x \in \{a, b\}$, and $\vec{w}_x$ is a

[21] I confirm this fact in Monte Carlo simulations in both logistic and OLS (not presented).
[22] To identify people who take longer, I regress response time on question and teacher indicators. The composite of the residual plus the teacher fixed effect reflects the average residualized time that the teacher took to respond to questions. The method identifies people who systematically take longer and shorter durations when rendering a decision to a given question. The only systematic association between taking longer and preferences appears to be that those taking longer express stronger preferences for defined contributions plans over defined benefits ($p < 0.001$).
[23] I use data I collected which can be found in Johnston (2024). Aldine ISD also provided several records Aldine (2020).

vector describing the working conditions, including contract features like the time-to-tenure. The data contain one row for each option a respondent considers, so each choice produces two rows in the data: one row for the option selected and another for the option forgone. For simplicity in the baseline model, I assume utility is additively separable in attributes, but the exercises produce very similar results when incorporating robustly significant interaction terms.[24]

Offers are indexed by j, and there is a finite set of them $j = 1, \ldots$,J. Each choice is made from a menu (choice set) indexed by s. Each choice set presents two offers, and each offer is characterized by a vector of K attributes: $X_{js} = [X_{js1}, \ldots, X_{jsK}]$. To estimate the influence of each factor, I use conditional logistic models as well as linear-probability models, regressing a respondent choice indicator on a vector of characteristics while conditioning on choice-set fixed effects to account for the options available to the teacher in each choice set:

$$u_{ijs} = X'_{js}\beta + \alpha_s + \varepsilon_{ijs} \quad (1)$$

Here, u_{ijs} represents an indicator for whether teacher i selected option j from choice set s and the coefficients capture how each attribute affects choice. Parts-worth utilities are denoted β and characteristics of alternative j are given by X_{js}. To aid interpretation in the main table, I convert parts-worth estimates into willingness-to-pay (WTP) by dividing each coefficient by the salary coefficient and multiplying by $1,000 (the unit of the salary variable). In the main analysis, the linear-probability model (which assumes the errors are uniformly distributed) is marginally better in explaining and predicting choice than the conditional logit (which assumes the errors are distributed extreme value). The standard errors are clustered by respondent to account for persistence in preferences across questions. Summary statistics for the attributes are presented in table 1, and a demographic description of teachers is presented in online Appendix table 2.

The Setting

[24] I assume additive separability for simplicity, but I test the importance of including interaction terms for the optimization exercise. The only robustly significant interaction term is that interacting salary and salary growth. (I estimate a saturated model with all the main effects and all the possible interaction effects. Just five of the 37 coefficients are significant. The adjusted R-squared increases by 0.006 from 0.192 to 0.198 when saturating with interaction terms. I include the significant variables in the main regression where all but one are no longer marginally significant.) I re-estimate models that include this interaction and re-calculate the welfare-maximal bundle. The resulting optima are virtually unchanged, suggesting the baseline model captures the systematic variation.

Aldine Independent School District has 70,000 students in Houston, Texas, with an annual budget of $700 million dollars (USDOE, 2016; NCES, 2019). Over three-quarters of students are eligible for free school meals (77 percent), which places the district at the 92nd percentile of student poverty among districts in Texas (calculation from data provided by TEA 2018, ESIS 2019). At the time the survey was delivered, the district had 4,358 full-time teachers who were invited to the district's yearly survey which, in 2016, included my choice experiment. The average teacher in the district has nine years of experience; 30 percent have advanced degrees. Just over two-thirds are female; the plurality are black (37 percent), and the remaining are white (28 percent) and Hispanic (21 percent) (online Appendix table 2). As is always the case whenever examining a single district, the results we find here may not generalize to other settings and are most likely to generalize to other urban, American settings.

III. Teacher Preferences for Compensation and Other Factors

The main preference estimates are presented in figures 2–4 and table 2. The figures visualize the results nonparametrically by showing estimates of model (1) with bins of each attribute, making it easy to gauge the response function of choice with respect to various offer characteristics. In table 2, I use the continuous variables to present part-worth utility βs and translate them into an interpretable (average) willingness-to-pay (WTP) for each trait; the left three columns represent estimates from linear probability models, and the right three represent estimates from conditional logistic models estimated with maximum likelihood. All estimates are harmonized across the three decks using subjects' responses to the salary feature.[25] Columns (3) and (6) represent a money metric, conveying how much teachers value a unit of that feature in terms of a permanent salary increase. These are the first direct estimates of teacher WTP for several attributes including elements of compensation structure, class size, contract attributes (time-to-tenure, review frequency), commuting time, and principal support.

Teachers value $1,000 of district subsidies for insurance equal to $970 in salary increases, suggesting the marginal utility is close to the marginal cost. (These two forms of compensation receive the same tax treatment: employer-paid premiums are exempt from federal income tax as are employee contributions (Brookings 2016)). An additional one-percent increase in salary

[25] Specifically, each coefficient in Deck 2 is multiplied by $\beta_{salary}^{Deck1}/\beta_{salary}^{Deck2}$, so estimates across decks are in directly comparable units. Each coefficient in Deck 3 is multiplied by $\beta_{salary}^{Deck1}/\beta_{salary}^{Deck3}$.

growth is valued equivalent to a permanent $2,270 increase in salary. This suggests that the average teacher expects to remain in teaching for at least six years, since only then does the total present value of an additional 1 percent growth exceed the total present value of $1,000 higher in starting salary using a five-percent discount rate.

Moving to a defined-contribution (DC) retirement plan from a traditional pension increases teacher utility equal to a salary increase of $907, presumably because DC plans are portable and insulated from perceived political risk. Prior work finds that public-sector workers are concerned about the future of their pensions because of underfunding (Ehrenberg 1980; Smith 1981; Inman 1982). Teachers value an additional ten-point replacement rate in pension equivalent to a $1,730 salary increase, somewhat less than its cost of $2,870 per year (see also Fitzpatrick, 2015). I use the tradeoff teachers are willing to make between higher salary today and higher retirement benefits in the future to calculate their intertemporal substitution parameter, δ—the discount factor. Teachers value a 1 percent increase in their retirement replacement the equivalent of a $173 starting-salary increase, which would increase their yearly retirement benefit by $840 under the current salary schedule after 30 years when teachers become eligible for retirement. The implied discount factor is 0.949 (solving for delta, $840 \times \delta^{30} = 173$), a value that closely matches the empirical literature estimating discount factors (Best et al. 2018; Ericson and Laibson 2018).[26] This supports the view that teachers respond realistically to the survey's hypothetical questions.

Teachers value performance pay but are averse to being evaluated only on the basis of value-added measures. An additional $1,000 in performance pay to the top quarter of teachers costs $250 per teacher ($1,000 × 1/4). On average, teachers value a thousand dollars in merit awards available at $346, a third more than the cost. The preference for performance pay in excess of its expected value may reflect its perceived nonpecuniary effects on the work environment.[27] Having rewards based solely on value-added measures is the equivalent of reducing a salary by $910. It is possible that teachers prefer Danielson scores because they can be influenced more easily than VA (Phipps 2018). While a teacher can prepare for a small

[26] The WTP for retirement income by new teachers implies a δ of 0.939.

[27] The district paid a consulting firm to calculate "student growth percentile" (SGP) measures. These scores compare each student's growth to that of students with similar initial scores. While the district ensured teachers received their Danielson scores (teachers were required to sign they had been informed of their Danielson score), the district made no use of SGP and there was no dissemination of SGP. SGP is correlated with VA (R-squared=0.121), but not highly predictive (44.6 percent of those with above average SGP have below average VA).

number of scheduled observations, success in value-added models (VA) requires continuous effort. Alternatively, teachers might prefer an objective measure to an observation score that could be permeated by bias, subjectivity, or favoritism. In the end-of-survey questions I ask a few more questions and learn that teachers also prefer a tandem evaluation over being evaluated by observation scores alone. What this implies is that teachers prefer having multiple, independent measures enter their evaluation. I test whether teachers' aversion to rewards based only on VA differs by whether the teacher has a relatively low VA compared to her Danielson score. Preferences do not differ by relative strength on VA or Danielson, suggesting that teacher preferences for evaluation are not strategic.

The presented job offers vary how many years teachers are evaluated before they are granted a permanent contract similar to tenure. Reducing the probationary period by one year (when it normally takes three years to receive permanent status) is valued at $470. The district also has regular review periods in which a teacher's performance is assessed once she has permanent status. More frequent reviews impose no disutility, suggesting they are not searching or demanding. An additional ten-minute commute is equivalent to a salary reduction of $530, implying that teachers are willing to commute for $9 per hour, half a teacher's hourly wage ($19). This is exactly consistent with the literature on willingness to commute (Small 2012; Mas and Pallais 2017).

Reducing class size by one student raises teacher utility the same as a $595 salary increase (1.2 percent of starting salary). Translating estimates of the effects of class size and compensation on teacher attrition, we can infer WTP from previous studies for comparison. Non-experimental estimates from Mont and Rees (1996) suggest that a teacher would give up 3 percent of her salary to reduce class size by one student; Feng (2005) finds no relationship between class size and teacher turnover, implying no preference for smaller classes. Teachers value an additional hour of teaching assistance each week at $260—less than the cost of hiring someone to provide assistance at minimum wage. This weak preference is possibly related to the costly nature of transferring tasks (Goldin 2014). The WTP for the first few hours of help is high, implying a low level of support would be a cost-effective means to improve teacher welfare. The third deck varies student attributes and school-leadership characteristics. Teachers prefer schools with higher-achieving students and fewer children in poverty, consistent with Antos and Rosen (1975).

The most predictive attribute in any deck is whether the principal is "supportive" or "hands-off" with disruptive students. Having a supportive principal provides utility equivalent to a permanent $8,700 increase in salary. The importance of this factor is so large that a supportive principal in the lowest-utility setting presented is preferred to a hands-off principal in the highest-utility one. To understand how teachers interpreted having a "supportive" or "hands-off" principal with disruptive students, I contacted a random sample of respondents, who indicated that a supportive principal meets with disruptive students, supports the teacher in enforcing discipline, and sides with the teacher in disputes over discipline with parents.

An important question is whether supportive principals reduce teacher aversion to working in hard-to-staff schools. I estimate models where achievement and poverty share are interacted with the supportive-principal indicator. Supportive principals erase 90 percent of the aversion to working in a low-achieving school and 85 percent of the disutility associated with teaching in a high-poverty setting (table 3). This suggests that disruptive students are perceived as costly by teachers and that principal support is effective at mitigating those costs.

IV. Using Compensation to Affect Selection

Whether compensation and working conditions can affect selection depends on whether excellent teachers have distinctive preferences. High-quality teachers may have weaker aversion to long probationary periods since they worry less about dismissal; they may have stronger preferences for small classes if they place a higher value on individual attention; high-quality teachers may place a greater value on starting salaries if they have more competitive outside options; and excellent instructors may have a distinct appreciation for generous pensions if they are more committed to a long career in teaching. It is also important to know whether highly effective teachers have different preferences for working conditions that are not affected by policy (like student demographics) to understand the payment levels needed to retain excellent teachers in hard-to-staff schools.

To measure teacher performance, I estimate value-added models (VA) from student achievement data and use Danielson observation scores. The student achievement data contain test scores for each student in each year they are tested with links to the student's teacher in grades 3–8 for years 2011 through 2016. I estimate VAs while controlling for a cubic polynomial of past scores, and I adjust for sampling error following Kane and Staiger (2008).[28] The VA model implies that a 1 standard deviation (SD) increase in teacher VA raises student test scores by 0.16 SD in math and 0.11 SD in reading, similar to Chetty et al. (2014a). The VA used in the primary analysis is the teacher's math value-added, and the results are similar when using reading VA.

Since students are not tested in all grades, there are records to estimate value-added for about half of teachers. To provide a measure of quality that covers a broader group of teachers, I supplement with Danielson observation scores generated by principal evaluations. I sum each teacher's four Danielson scores (one for each of four categories: *planning and preparation, classroom environment, instruction,* and *professional responsibilities*). VA and Danielson are significantly correlated, but also have a significant independent component (R correlations between Danielson ratings and VA range from 0.153 to 0.212 depending on the value-add model). Together, the VA index and the Danielson index provide a quality measure for about 80 percent of respondents.

To test whether preferences vary by teacher quality, I interact each of the attributes from table 2 with the quality measure in a model of teacher choice. I condition on experience dummies that indicate having exactly *n* years of experience to account for the fact that more experienced teachers may systematically have higher value-added and have different preferences (table 4).

The most highly rated teachers have similar preferences to their colleagues for almost all school attributes (table 4 and online Appendix tables 3 and 4). High-quality teachers do not, for instance, have a stronger preference for more generous pensions, higher salary, or high-performing students. In terms of work setting characteristics that policymakers can influence, effective teachers have the same preferences as other teachers with regards to class size, salary, income growth, insurance subsidies, retirement benefits, and supportive principals.

The only way in which high-performing teachers systematically differ is their preferences for offers including performance pay (table 4 and figure 5). A teacher in the bottom decile values a $1,000 merit reward as equivalent to a $83 salary increase. Teachers in the top decile value the same merit program as equivalent to a $510 salary increase (the interaction $p < 0.001$).[29] If teachers

[28] Additional details about VA estimation can be found in online Appendix B.

[29] In the district, teachers are informed their VA measure and Danielson score each year, so they know their placement in the distribution. Why then do low-performers have some preference for offers containing performance pay. Potentially,

entertained two comparable offers, a high-performing teacher (top decile) is 23 percent more likely than a bottom-decile one to select the offer providing an additional $3,000 in merit pay per year. Over time, this preference generates positive selection in retention where performance pay is implemented. Since the best performers receive increased compensation, the probability of attrition is reduced relative to teachers with lower performance. Whether performance pay can generate favorable selection *into* teaching is not clear from this examination. Performance pay may not affect selection on entry if prospective teachers do not have private information about their ability to teach. It bears mention that the district does not calculate or distribute estimates of VA to teachers. Since high VA teachers prefer performance pay, it implies teachers have private information about their output.

The relationship between teacher quality and preferences for performance pay is illustrated in figure 5. Deciles 2 through 7 express differential preferences that are very close to zero. Teachers in deciles 8, 9, and 10, however, have significantly stronger preferences for performance pay than lowest-decile teachers, being 3.8 percentage points more likely to select an offer containing an additional $1,000 in performance pay to the top quarter.

In future work, it may be fruitful to examine whether there are differential preferences for other attributes including dismissal rules and the quality of coworkers.

The Preferences of Marginal Teachers

An important dimension of heterogeneity is whether marginal teachers (those close to indifference between remaining in or exiting the profession) have similar preferences to their inframarginal peers. For marginal teachers, changes in the compensation structure are more likely to affect their labor-supply decision. I incorporate information on which teachers who took the survey in 2016 left the district by 2018 and interact an indicator for leaving with each attribute while controlling for experience dummies and experience bins interacted with each attribute. Marginal teachers have statistically identical preferences for compensation structure and student characteristics. Of the 18 attributes in the study, teachers who leave the profession have statistically different preferences in two, both at the five-percent level. Leavers have weaker aversion to large classes and stronger interest in having teaching aids. In the compensation attributes we explore in the next section, leavers have statistically identical

---

low-rated teachers believe they can improve their instruction to benefit directly from the incentive, or low-rated teachers believe the incentive would improve the professional environment.

preferences (online Appendix tables 5–7). What this suggests is that attrition is largely unrelated to preference heterogeneity for working conditions. What differs between leavers and stayers is their taste for *teaching*.

To examine whether the preferences of marginal teachers differ on entry, I survey 1,193 college students at the University of Houston—a large public university near Aldine ISD. Students are asked to report how likely they are to enter teaching (on a Likert scale from "*I would never consider teaching*" to "*I plan to be a teacher*"). I ask respondents to imagine that, regardless of their interest in teaching, they decided to become a teacher for one year. They then respond to the same choice experiment used in the district to elicit their preferences for compensation structure and working conditions. It bears mention that college students may not have the same incentives for truth-telling since the survey is not sent by their employer or used to inform their employer's policy; the survey is, however, distributed by the career services office at University of Houston, and respondents may have believed that their responses would have helped tailor their services to meet their needs or form recommendations.

What is of interest here is whether those planning on teaching have similar preferences for compensation structure as those who are on the margin. If those that could be induced into teaching have different preferences, then improving the appeal of teaching would change the composition of preferences in the teaching profession making a counterfactual exercise especially challenging. Comparing the preferences of those set on teaching with those seriously considering it reveals no difference in preferences. The significance in the interacted terms (attributes interacted with "teaching propensity") is null in each model. Even when including the full spectrum of interest in teaching, preferences differs little-to-none along the interest-in-teaching index (online Appendix tables 11–13). The joint significance of attributes interacted with the teacher-propensity index is not statistically significant in the compensation deck. The adjusted R-squared in the base model is 0.174. When the attributes are interacted with interest-in-teaching, the adjusted R-squared is still 0.174. In the working condition deck, some differences emerge. Areas in which inframarginal teachers differ from other respondents are in attributes where those investigating the profession would have a clearer view. For instance, those who plan on teaching have a stronger aversion to large classes and a stronger preference for supportive principals than those who do not intend to teach. Even here, however, the explanatory power of the model (as measured by adjusted R-squared) is improved by just 0.002 when including interest-in-teaching interactions in the working condition deck.

This exercise suggests that tastes for compensation structure are essentially uniform along the distribution of interest in teaching, which has a few implications. First, it implies that the preferences uncovered in the experiment will generalize to teachers on the entry and exit margins. Second, it implies that if compensation were made more attractive, any differences in selection into the profession will have little impact on the composition of teacher preferences. Importantly, this suggests the stability of counterfactual policy experiments since new teachers, induced into the profession by reform, won't have distinct preferences that would imply a different optimal compensation structure. The uniformity of preferences along the interest-in-teaching dimension, moreover, implies that it is tastes for teaching that drives entry and exit, not different preferences for compensation structure among teachers.

Preference Heterogeneity in Other Dimensions

I examine several other dimensions of preference heterogeneity. To study how preferences differ by experience level, I divide teachers into four experience quartiles: novices, who have 0–1 years of experience; new teachers, who have 2–6 years of experience; experienced teachers, who have 7–14 years of experience; and veterans, who have 15 or more years of experience. I then interact dummies for "new," "experienced," and "veteran" with each attribute and estimate models like equation (1). The main estimate provides the preferences of novice teachers (the omitted category). The interaction coefficients show the preference differential from novice teachers for each experience category. Results are presented in online Appendix tables 14–16.

More experienced teachers have weaker preferences for higher salary and stronger preferences for more generous retirement plans. In working conditions, preferences are similar to those of novices in time-to-tenure, term length, and commute time, but older teachers have a higher tolerance for larger classes and a stronger demand for teaching assistance. Senior teachers also have stronger preferences in favor of high-achieving students than their less experienced colleagues. Novice, new, and experienced teachers have similar preferences for having a "supportive" principal, but veteran teachers place an additional premium on it. A district could attempt to retain veteran teachers by providing compensation options that suited the preferences of established teachers.

Black-white and male-female achievement gaps may partly be the byproduct of teacher demographics, based on positive estimated effects of having teachers of the same race and sex (Goldhaber et al. 2019). Some policymakers would like to attract and retain teachers to mirror the demographics of the student population.

To study how preferences differ by sex, I interact male indicators with each attribute (see online Appendix tables 17–19). Men have stronger preferences for salary than women and are more averse to high-deductible health plans, consistent with women being more likely to receive health insurance through a spouse. Like veteran teachers, men are more willing to teach large classes, and they place a lower value on assistance with grading. Men and women have similar preferences for student demographic characteristics, but men exhibit less demand for a supportive principal. Race heterogeneity is presented in online Appendix tables 20–22. Black teachers have weaker preferences for salary growth than white and Hispanic teachers. Black and Hispanic teachers have stronger preferences for performance pay than white teachers. Black teachers place higher value on a short tenure clock and less frequent reviews than white and Hispanic teachers. All three groups have similar preferences for commuting and assistance with grading.

While white and Hispanic teachers have precisely zero preference for student race, black teachers prefer student bodies that have a higher minority share, mirroring Antos and Rosen (1975). While everyone has strong preferences for a supportive principal, black and Hispanic teachers value supportive principals 8–12 percent less than white teachers. That both male and minority teachers have weaker preferences for principal support suggests they either experience lower costs of classroom disruption or enjoy additional social capital with disruptive students. Finally, I test whether preferences differ by grade level. In general, teachers in elementary schools, middle schools, and high schools have similar preferences for compensation, student attributes, principal affect, commuting, and assistance. Middle and high school teachers, however, express less aversion to large classes and stronger aversion to longer tenuring periods than elementary-school teachers (see online Appendix tables 23–25).

V. Optimizing the Compensation Structure

Compensation Structure

What do preferences suggest about how schools should structure compensation? I consider three related objective functions schools might pursue, and I maximize them subject to the current budget constraint. First, I calculate the compensation structure that maximizes teacher *utility*. Second, I calculate the compensation structure that maximizes teacher *retention*. Third, I calculate the compensation structure that maximizes student *achievement.* These three objectives nest one another: teacher welfare affects retention, and teacher retention affects student achievement.

In practice it is difficult to know what objective function school districts pursue. Economic models of nonprofit institutions usually entail the maximization of "prestige" or sometimes the quantity of services delivered (Newhouse 1970; Johnston and Johnston 2021). Schools are operated by elected boards, so they likely seek to maximize outcomes that appeal to voters—the most determined of which are parents and teachers. The objectives of these voters will vary substantially from maximizing achievement to maximizing compensation to maximizing the football win rate. Teachers' unions also have significant leverage to advance their objectives in schools—where models typically assume unions maximize the total compensation bill (Freeman and Medoff, 1984). A social planner wanting to maximize social welfare would likely be interested in maximizing human capital given available resources, since achievement profoundly shapes lifetime production, consumption, and welfare. Though we don't know the actual objective function of districts, it is useful to compare what districts do to a benchmark of what they would do if they were maximizing student achievement (e.g., Hoxby 1996, Walters 2018).

All simulations are based on the same model of teacher utility estimated from my field experiment. In using estimated utility, I implicitly assume that incoming teachers have similar preferences, and a similar quality distribution, to those already in the district. The first assumption is supported by our examination of preferences along the propensity-to-enter-teaching index. The second assumption produces a conservative estimate of achievement effects if performance pay encourages positive selection on entry (Jones and Hartney 2017). The optima in some exercises fall outside of the experimental range which can be a concern. Since preferences are primitives, the extrapolations resulting from optimization tend to perform well in predicting out-of-sample effects (Todd and Wolpin 2006).

The Optimization Problem

I begin by specifying the objective functions I maximize. The first is an objective to maximize teacher utility. This may be the goal of a district with a decisive union presence that represents the preferences of its members. To simulate the optimal pay structure for teacher utility, I estimate teacher utility models that allow for diminishing marginal returns by including a squared term of relevant non-binary features including salary growth, class size, performance pay, and the replacement rate in retirement. In theory, we expect there to be diminishing returns to each continuous attribute, which is confirmed in estimation; each attribute has a negative second-derivative coefficient and most are statistically significant at conventional levels. (Only the coefficient on growth-squared is insignificant at conventional levels, having a p-value of 0.14 in a one-sided test.) I include starting pay as a linear numeraire (for estimates, see online Appendix tables 26 and 27 or footnote 26).

$$(2) \quad U_a = (\widehat{\beta_1^1}\, S_a + \widehat{\beta_2^1} G_a + \widehat{\beta_3^1} G_a^2 + \widehat{\beta_4^1} P_a + \widehat{\beta_5^1} P_a^2 + \widehat{\beta_6^1} M_a + \widehat{\beta_7^1} R_a + \widehat{\beta_8^1} R_a^2 + \widehat{\beta_9^1} D_a + \widehat{\beta_{10}^1} H_a)/\widehat{\beta_1^1} + (\widehat{\beta_2^2} C_a + \widehat{\beta_3^2} C_a^2)/\widehat{\beta_1^2}$$

Here, the utility of an allocated bundle a is a function of starting salary (*S*), the growth rate (*G*), performance pay (*P*), the basis of performance pay (*M*), the retirement replacement rate (*R*), the retirement plan type (*D*), health insurance subsidies (*H*), and class size (*C*).[30] Without allowing for diminishing marginal returns, the results could degenerate to a corner solution in which all compensation loads into the attribute with the highest average utility per dollar. The parameter β_x^y refers to the coefficient on variable number x displayed in deck y. To aid interpretation, I convert *utility* into a money-metric equivalent by dividing each of the coefficients by the beta on starting salary in the same deck (β_1^1 in deck 1 and β_1^2 in deck 2). The resulting scale of utility is its money-metric equivalent in 1,000s of dollars. The calculation maximizes the average utility of respondents (Hoxby 1996; Figlio 2002).

[30] The values used in this equation are from online Appendix tables 29 and 30: $U_a = (0.0846S_a + 0.2225G_a - 0.0145G_a^2 + 0.1326P_a - 0.0386P_a^2 - 0.0767M_a + 0.0388R - 0.00024R^2 + 0.0767D + 0.0821H)/0.0846 + (0.0916C - 0.0029C^2)/0.0787$

Maximizing retention (in terms of mean teacher experience) is a related objective, and one that would naturally be of interest to schools since teacher experience is one of few reliable predictors of teacher performance (Wiswall 2013). Hendricks (2014) estimates the effect of salary on retention over the life cycle of a teacher. These estimates neatly link to my utility framework since utility is converted to salary-equivalent units. Because the estimates in Hendricks (2014) come from Texas, they likely generalize to my setting. Let λ_e denote the baseline retention rate for each experience level, *e*, and let η_e represent the change in retention rates for a percent change in salary which varies with experience. The retention probability at experience level *e* is calculated:

(3) $$r_{ea} = \lambda_e + \eta_e \times \Delta_{\mathrm{a}}$$

The Δ_{a} is the difference in utility between the compensation in Hendricks and the salary-equivalent utility of the bundle under consideration, U_a from equation (2), where the difference enters the model as a percent. The average tenure predicted using status quo compensation structure matches the district's actual average experience (9.0 years), suggesting the convergent reliability of the model. To determine the average tenure, I calculate the share of teachers remaining in each experience cell and simulate the equilibrium distribution of experience. The stock of teachers in experience cell *e*, is simply the number remaining from the prior experience cell, $e-1$, (where the stock persisting in year *e* is calculated $S_e = S_{e-1} \times r_{e-1}$). I normalize the shares so they sum to one and denote the distribution of these normalized shares $\mathbf{D}_{\mathrm{e}} = [D_1, D_2, \dots, D_{35}]$ where D_e states the share of teachers employed with experience level *e* in equilibrium. The object I maximizes is the *average* teacher experience produced by compensation allocation *a*, calculated:

(4) $$E(a) = \sum_{e=1}^{35} D_e(a) \times e$$

The last objective I consider is the maximization of student achievement. The model incorporates several well-studied features of the achievement production function.

- First, students learn more in smaller classes (Krueger 1999; Hoxby 2000; Cho, Glewwe, and Whitler 2012).

- Second, improving teacher welfare affects the retention probabilities in each experience cell; retention increases achievement since teacher experience has increasing, concave impacts on student learning (Papay and Kraft 2015).
- Third, performance pay improves achievement by eliciting additional effort (Lavy 2009; Imberman and Lovenheim 2015). Performance pay also produces positive selection in retention because high-VA teachers have stronger preferences for performance pay.

The achievement production function I maximize uses the average per-student impact of class size from domestic studies (Krueger 1999; Hoxby 2000; Cho, Glewwe, and Whitler 2012). The average estimate implies an additional 10 students per teacher reduces student gains by 0.012σ per year. To calculate the influence of experience on achievement, I calculate retention probabilities, simulate the equilibrium experience profile, and take the dot product of the experience distribution with the VA-gains-by-experience vector from Papay and Kraft (2015). For performance pay, we use an estimate from Imberman and Lovenheim (2015) that a thousand dollars in performance pay increases achievement via effort by 0.014 standard deviations. I use Imberman and Lovenheim because their estimate has the virtue of being from the same setting as my experiment.[31]

So that the resulting bundles are directly comparable to the status quo, each objective is maximized subject to the current budget constraint, which takes the form:

(5) $$\{\boldsymbol{S} \cdot \mathbf{D}_e \times (1 + \phi R + U) + T(t) + P/4 + H\}N < B$$

Here, $\boldsymbol{S}$ is the salary schedule implied by a starting salary S and a growth rate G. The cost implied by the dot product between the salary schedule and the equilibrium distribution of teacher experience is the equilibrium cost of salary. In order to provide a replacement rate R, the district has to pay a fraction of salary ϕR into retirement accounts. The cost parameter ϕ, therefore, reflects the needed contribution for a one-percent replacement in retirement. The

[31] The achievement function can be written: $A = \boldsymbol{D_e} \cdot \mathbf{VA}_e - 0.0012 \times C + 0.0136 \times P$, where $\boldsymbol{D_e}$ is the experience distribution of teachers from resource allocation a, $\boldsymbol{VA_e}$ is a vector of VA gains arising from experience e; C is the class size of the allocation; and P is the performance pay of the allocation. The estimate is potentially conservative because it is the effect of a 10 percent increase in exposure, and Imberman and Lovenheim find the achievement returns scale until 20 to 30 percent exposure.

cost parameter is 2.5 times greater for defined-benefit plans than for defined-contributions plans because it is more expensive to guarantee income. There is a budget cost to turnover (t) which reflects the cost of recruiting, onboarding, and training (Barnes, Crowe, and Shaefer 2007; Watlington et al. 2010). Retention is, therefore, cost-saving. Turnover is calculated by summing the departures calculated when simulating the experience distribution. Some small per-person costs, *U*, are required, which captures the cost of unemployment insurance, Medicare taxes, and workman's compensation. *P* is the performance pay provided to the top quarter of performers each year, and *H* is the health insurance subsidy provided to each worker. The number *N* is the quantity of teachers a principal would need to provide a class size C to a grade of 100 students, where teachers are perfectly divisible (*N* = 100/*C*). The search operates such that the total cost must be no more than the current personnel cost of educating 100 students, \$291,572 per year. The costs interact. For example, retirement replacement becomes more expensive as salary increases, and class-size reductions become more costly as total compensation rises.[32]

I constrain optimization to conform to some practical considerations. No unit of compensation can be negative. I've included class size as a way of testing whether smaller classes function as a cost-effective compensation provision to teachers or a cost-effective means of achievement promotion. In the main experiment, therefore, I constrain class size so that it cannot rise to greater than 30 students per class (the status quo is 28.7). I present optima that allow any class size in the online Appendix. Without the constraint on class size, the model pushes towards large classes (43 students per class when maximizing achievement) with highly paid teachers (starting salaries of \$110,700). Relaxing class-size constraints allows for additional student growth of 0.01σ per year. Performance pay is also constrained so that it can be no larger than \$5,000. This limit is somewhat beyond the experimental range I test and near the limit of the quasi-experimental range of the natural experiment I rely on (Imberman

---

[32] The budget constraint inequality is $\{\boldsymbol{S} \cdot \mathbf{D}_e \times (1 + ((1 - DB) \times 0.00220R) + (DB \times 0.00559R) + 0.015) + (11.891 \times t) + (0.25 \times P) + H\}N < \$291.571$. The salary bill is the dot product of the salary-schedule vector $\boldsymbol{S}$ and the experience distribution $\mathbf{D}_e$. A fraction of salary costs are set aside for each unit of replacement in retirement. It costs two and a half times more to fund replacement through a defined benefit vehicle (*DB*) that guarantees replacement. The district also pays an additional 1.5 percent of the salary bill to finance unemployment insurance and workman's compensation which I learned from their financials. When calculating the experience distribution, we sum the number of teachers that leave each year, *t*. Accounting studies have found that it costs \$11,891 to recruit, onboard, and train each new teacher. The part in brackets ({}) is the cost per teacher. I multiply the cost per teacher by the number of teachers, *N*, needed to provide class size *C*, to calculate the total cost. The total personnel cost of teaching 100 students is constrained to be no more than the current personnel cost, which in 2016 was \$291,571 per 100 students.

and Lovenheim, 2015). Without this constraint, achievement maximization recommends substantial allocations of performance pay that extrapolate well beyond the range tested by my experiment or Imberman and Lovenheim (2015). Constraints on starting salary, growth, and retirement replacement never bind. Binary attributes (defined contributions indicator and using-VA-only evaluations) are constrained to be within [0,1]. I provide additional details about the maximization exercise and cost calibration in online Appendices C and D. Results are presented in Table 5.

Compensation Structure to Maximize Teacher Utility

At the time of the survey, the district paid $50,000 in base salary, with a 1.8 percent average yearly increase in salary earnings. They provided no performance pay, had an average class size of 28.7 students, paid $3,960 in health-insurance subsidies, and it replaced 69 percent of a teacher's top earnings in retirement through a pension program after 30 years of service.

To maximize teacher utility subject to the current budget constraint, the school would pay 50 percent more in base salary ($75,655) and offer $1,477 in merit pay to the top quarter of teachers. These increases are financed by reduced expenditure elsewhere: slightly increased class size (4.5 percent), reductions in salary growth (from 1.8 percent growth to 0.0), and a reduced replacement rate in retirement (20 percent). Schools would also shift to defined-contributions retirement plans that are both less costly to districts and preferred by teachers—a Pareto improvement across agent types. In total, these changes increase teacher welfare by 21.6 percent, the equivalent of a $17,000 increase in annual salary—without spending additional money. Utility improvements are generated by salary increases (91.6%), the introduction of merit pay (5.0%), and shifting toward a defined-contributions retirement plan (3.4%).

I assess the influence of this compensation structure on other outcomes using the objective functions specified in the last section. Maximizing teacher utility would intuitively increase teacher retention and thereby raise average teacher experience by 21 percent in equilibrium. This reform also increases student achievement by 0.043σ each year, which comes from increased teacher experience (47%), induced effort from merit pay (46%), and improved retention of highly productive teachers (7%).

Moving to a defined-contributions plan may not be politically feasible in the current environment. To understand the optimal replacement rate without shifting to a DC retirement program, I re-calculate the optimal structure constraining the model to use a traditional pension. The calculation suggests an optimal replacement rate 55.5 percentage points (or 80 percent) lower than the status quo, owing to a higher salary (which makes replacement more expensive) and the expense of guaranteeing income.

Compensation Structure to Maximize Teacher Experience

Experience reliably predicts teacher effectiveness, and recent evidence suggests that teacher output improves throughout a teacher's career (Wiswall 2013; Papay and Kraft 2015). Whenever any teacher departs the profession, it opens a vacancy chain that causes an additional novice to be hired somewhere, reducing achievement on average.

The compensation structure that maximizes retention implies starting salary above the status quo ($66,688). The optimum targets higher compensation to teachers that already have experience with a positive salary growth rate of 1.3 percent. Like the utility-optimal bundle, the retention-optimal bundle offers performance bonuses of $1,487 for the top quarter of teachers each year. These increases are paid for with larger classes (4.5 percent) and 18 percent lower replacement rate in retirement. Lower replacement rates overstate the reduction in retirement income since the replacement rate applies to a higher final salary.[33]

The optimal structure for maximizing retention increases average teacher experience by 22 percent, raises the odds that a student has a veteran teacher by 33 percent, and reduces the chances they have a novice by 28 percent. When compared to the utility-maximizing bundle, the retention-optimal structure increases average teacher experience using a higher salary growth rate that improves the odds of retaining teachers who already have a stock of experience. The changes produce a 0.044σ increase in student achievement each year, an improvement that arises from an increase in teacher experience (47%), an increase in teacher effort from performance pay (46%), and positive selection in retention (7%).

[33] The replacement rate for DB is a third as large than the status quo, but the resulting retirement annuity is half as large owing to the higher salary replaced. I implement an alternative model which excludes retirement preferences from utility and uses retention effects from pensions estimated in Costrell and McGee (2010), who estimates the influence of pension wealth accumulation on attrition. Pensions benefits are backloaded, so they produce a strong pull for teachers nearing retirement when pension benefits spike while doing little to retain younger teachers. These simulations suggest that defined contributions plans increase teacher tenure more than pensions, consistent with regression-discontinuity evidence in Goda, Jones, and Manchester (2017). The logic is twofold: teachers prefer defined contributions, and the marginal accretion of retirement wealth is larger for more teachers in DC plans than in DB.

Compensation Structure to Maximize Student Achievement

Improving teacher welfare does not necessarily increase achievement (for example, de Ree et al. 2018). The reform that maximizes achievement would include higher base pay than the status quo ($66,771), a modest rate of salary growth (1.3 percent growth rate), $5,000 payments in performance pay, and a slightly larger class size (4.5 percent). The resulting achievement-optimal bundle reduces the replacement rate by 18 percent in retirement, relative to the status quo, while moving to a defined-contributions retirement plan. This structure increases student achievement in equilibrium by 0.091σ per year while also improving teacher welfare by 7 percent at the same time. The achievement gains largely come from effort induced by merit pay (75%), with more modest components coming from more experienced teachers (10%) and increased retention of high-caliber teachers (15%). The gains coming from altered retention (twenty-five percent of the total effect) take form over time. About 20 percent of those gains are realized in the first five years; 80 percent of the potential gains are realized in the first 20 years; and the whole effect is realized in 35 years. The bulk of the gains, however, come from effort which would be active immediately.

The model predicts partial-equilibrium effects where a single district alters its compensation structure. How would the effect scale in general equilibrium if all districts adopt similar reforms to maximize achievement? The key is to understand the degree to which local compensation-induced retentions prevent (i) transfers to other school districts or (ii) departures from the teaching profession as a whole. Retaining teachers who would have transferred to other districts comes at the expense of other districts—and thus those partial equilibrium gains would disappear in general equilibrium. Preventing departures from the teaching profession do not come at the expense of other school districts, so those gains remain in general equilibrium.

I estimate the effect of compensation on (i) district transfers and (ii) professional departure using administrative staffing records from Texas and a triple-difference design following Hendricks (2014). This approach leverages variation within a district across experience cells when districts change their salary schedules. As an example, the identifying variation might come from a district that increased starting salaries while maintaining its salary levels for veteran teachers, or the reverse. Using these fine-grained comparisons, I find that salary increases have a small, insignificant effect preventing transfers, but a large effect preventing early retirement.[34] This implies districts have significant market power as employers, which is consistent with their typically employing more than 90% of teachers in an area. To extend my partial-equilibrium results to a general equilibrium context, I assume that the direct effects of class size and effort scale in general equilibrium but not necessarily all the benefits from retention where salary-induced retentions come at the expense of other districts. The results imply that the general equilibrium effect on achievement would be just 1.6 percent smaller than the partial equilibrium one. More details on this approach and its results are provided in online Appendix E.[35]

Across objectives, the maximization exercises imply an increase in salary and merit pay and a reduction in the replacement rate while moving towards defined-contributions retirement programs would improve outcomes. The achievement-maximizing structure recommends a level of performance pay that roughly mirrors the share of compensation private sector workers receive in bonuses, about 2 percent of total compensation (BLS 2018).

The experimental variation reveals the preferences for a given group of workers. The estimates indicate whether district compensation is distorted from its own optimal based on teachers already there. It is striking that—even among a group of teachers selecting into the district—the status quo compensation structure does not appear to reflect either teacher preferences or a structure that maximizes retention or achievement. If the calculated optimal structures were similar to the district's practice, we might suspect that it reflects endogenous sorting into the district. The fact that optimal structure diverges from the status quo among an endogenously selected group suggests that working conditions and compensation structure are structured especially poorly.

VI. Discussion and Conclusion

The district's compensation scheme does not conform to goals of maximizing teacher welfare, teacher retention, or student achievement. It may be that political constraints or bargaining distort compensation structure from an optimum. Since unions are typically led by

[34] It appears that the local market power of school districts (who set compensation at dozens or hundreds of nearby schools) make the primary margin of choice whether teachers continue teaching rather than which district to teach in.
[35] If other professions adjusted compensation in response to reforms in teaching, like Willen (2021), the general equilibrium effects would be somewhat smaller.

older, veteran teachers, they may bargain for compensation structures that reflect their private preferences, especially if the costs are shrouded to voters (Glaeser and Ponzetto 2014).[36]

To evaluate the generalizability of the recommendations for optimal structure, I compare the district's compensation structure to the rest of the state and country.[37] One of the consistent findings from the maximization exercise is that the district can improve teacher welfare, retention, and student achievement by increasing salary expenditures as a fraction of total compensation. If the district has low salary share compared to other districts, it may simply fall on the high side of a distribution that is centered on what is optimal. That is not the case. Two-thirds of school districts have salary shares below the district we examine; when weighting by the number of teachers in a district, we learn that 90 percent of teachers are in school districts with salary shares lower than the district. Since the district appears to underinvest in salary, the many districts who invest even less are likely also underinvesting.

The results highlight several avenues for future work. Work that examines *entry* and *exit* in the teaching profession would provide a more tailored equilibrium examination of teacher sorting. To examine entry, analysts could measure how policy variables affect career choice for teaching among college students using a similar hypothetical choice design. This would further illuminate how to efficiently draw able individuals into the teaching profession. To examine exit, researchers might field a similar set of questions as those I've presented with options to leave the profession either for home production or their preferred alternative career. Because of the potential benefits of separating equilibria, designs that examine whether excellent teachers have distinct preferences for colleague quality, dismissal risk, or other attributes may afford policymakers with additional tools to (differentially) recruit and retain excellent instructors. And, considering the apparent importance of principals for teacher welfare, a deeper examination of principal influence would pay dividends.

In this study, I use a choice experiment to measure teacher preferences for compensation and working conditions. This approach has several advantages. First, the variation in attributes I study is credibly exogenous. It is not variation generated by endogenous political or market processes. Second, the design allows me to introduce independent variation in important

[36] Indeed, I find that teachers value more generous retirement plans the more senior they are, and the relationship is strictly monotonic for bins of teacher experience.

[37] Compared to teachers in other districts, teachers in the district receive total compensations at the 55th percentile in Texas and the 24th percentile across the country.

attributes that do not vary in the real world. In practice, competing schools offer the same compensation and contracts for a variety of explanations including market concentration, pattern bargaining, and public regulations (Biasi 2021). Third, in addition to the fact that researchers find a high degree of realism in response to hypothetical choice, the use of my experiment to inform the district's new compensation ballasts the incentives for teachers to provide truthful responses.

This study demonstrates how teachers value a wide variety of compensation vehicles, contract types, and working conditions. Most of these estimates are novel in themselves. Using real performance measures from administrative data, I test whether high-performing teachers have distinctive preferences that can be used to shape selection. Consistent with theory, preference differences between high-performers and their colleagues imply that performance pay can alter selection over time (Lazear 2000). Other compensation, contract, and working-condition attributes provide little scope for enhancing selection.

Using a model of teacher utility and estimates on selection, I model how schools would structure compensation and costly working conditions to achieve various objectives. What's surprising is that the optimal structures under a variety of objectives are substantially different from the status quo, and these simulated bundles are each more similar to one another than any are to current structure. Each implies a higher investment in salaries and performance pay to teachers at the expense of generous defined-benefits retirement programs. In each, both achievement and teacher welfare are simultaneously improved.

References

Aldine Independent School District. *Confidential Administrative Staffing Records: Retention, Evaluations, Assignments and Demographics*. Aldine ISD District Offices. (2016-2020). Unpublished data. Accessed 2020-10-10.

Altonji, Joseph G., and Richard K. Mansfield. “The Role of Family, School, and Community Characteristics in Inequality in Education and Labor Market Outcomes.” *Whither Opportunity* (2011): 339-358.

Antos, Joseph R., and Sherwin Rosen. “Discrimination in the Market for Public School Teachers.” *Journal of Econometrics* 3, no. 2 (1975): 123-150.

Ballou, Dale. “Do Public Schools Hire the Best Applicants?” *The Quarterly Journal of Economics* 111, no. 1 (1996): 97-133.

Ballou, Dale, and Michael Podgursky. “Recruiting Smarter Teachers.” *Journal of Human Resources* (1995): 326-338.

Ballou, Dale, and Michael John Podgursky. *Teacher Pay and Teacher Quality*. WE Upjohn Institute, 1997.

Barnes, Gary, Edward Crowe, and Benjamin Schaefer. “The Cost of Teacher Turnover in Five School Districts: A Pilot Study.” *National Commission on Teaching and America's Future* (2007).

Baron, Jason E. “Union Reform, Performance Pay, and New Teacher Supply: Evidence from Wisconsin’s Act 10.” Working paper (2020).

Barro, Robert J. “Education and economic growth.” *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-being* 79 (2001): 13-41.

Bates, Michael D., Michael Dinerstein, Andrew C. Johnston, and Isaac Sorkin. *Teacher Labor Market Equilibrium and Student Achievement*. No. w29728. National Bureau of Economic Research, 2022.

Best, Michael Carlos, James Cloyne, Ethan Ilzetzki, and Henrik Kleven. “Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches.” No. w24948. National Bureau of Economic Research, 2018.

Biasi, Barbara. “Higher Salaries or Higher Pensions? Inferring Preferences from Teachers' Retirement Behavior.” *Working Paper* (2019).

Biasi, Barbara. “The Labor Market for Teachers under Different Pay Schemes.” *American Economic Journal: Economic Policy* 13, no. 3 (2021): 63–102.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. “Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Disentangling the Preferences of Teachers and Employers.” *Journal of Labor Economics* 31, no. 1 (2013): 83-117.

Bratsberg, Bernt, and Ole Rogeberg. "Flynn Effect and its Reversal Are Both Environmentally Caused." *Proceedings of the National Academy of Sciences* 115, no. 26 (2018): 6674-6678.

Brookings Institute (2016). “Briefing Book: Key Elements of the U.S. Tax System.” Tax Policy Center. https://www.taxpolicycenter.org/briefing-book/how-does-tax-exclusion-employer-sponsored-health-insurance-work (accessed March 23, 2020).

Brown, Christina, and Tahir Andrabi. “Inducing Positive Sorting through Performance Pay: Experimental Evidence from Pakistani Schools.” Working paper (2020).

Bureau of Labor Statistics. *News Release Bureau of Labor Statistics: Employer Costs for Employee Compensation.* December, 2018, https://www.bls.gov/news.release/pdf/ecec.pdf.

Bureau of Labor Statistics. *Retirement Costs for Defined Benefit Plans Higher than for Defined Contribution Plans*. December, 2012, www.bls.gov/opub/btn/volume-1/pdf/retirement-costs-for-defined-benefit-plans-higher-than-for-defined-contribution-plans.pdf.

Burgess, Simon, Ellen Greaves, and Richard Murphy. "Deregulating teacher labor markets." *Economics of Education Review* 88 (2022): 102253.

Card, David, and Alan B. Krueger. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100, no. 1 (1992): 1-40.

Carson, Richard T., Theodore Groves, and Mark J. Machina, (2000), "Incentive and Informational Properties of Preference Questions," Paper presented at the Kobe Conference on Theory and Application of Environmental Valuation, Kobe: Kobe University, January.

Carson, R. T., Theodore Groves, John List, and M. Machina. "Probabilistic Influence and Supplemental Benefits: A Field Test of the Two Key Assumptions behind Using Stated Preferences." *Unpublished manuscript* (2006).

Chetty, Raj, Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *The Quarterly Journal of Economics*, *126*(4), 1593-1660.

Chetty, Raj, Friedman, J. N., & Rockoff, J. E. (2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *The American Economic Review*, *104*(9), 2593-2632.

Chetty, Raj, Friedman, J. N., & Rockoff, J. E. (2014). "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *The American Economic Review*, *104*(9), 2633-2679.

Cho, Hyunkuk, Paul Glewwe, and Melissa Whitler. "Do Reductions in Class Size Raise Students' Test Scores?" *Economics of Education Review* 31, no. 3 (2012): 77-95.

Chingos, Matthew M., and Martin R. West. "Do More Effective Teachers Earn More outside the Classroom?" *Education Finance and Policy,* 7, no. 1 (2012): 8-43.

Clemens, Jeffrey, and David M. Cutler. "Who Pays for Public Employee Health Costs?" *Journal of Health Economics* 38 (2014): 65-76.

Correa, Juan A., Francisco Parro, and Loreto Reyes. "Self-Selection in the Market of Teachers." *Applied Economics* 47, no. 13 (2015): 1331-1349.

Costrell, Robert M., and Josh B. McGee. "Teacher Pension Incentives, Retirement Behavior, and Potential for Reform in Arkansas." *Education Finance and Policy* 5, no. 4 (2010): 492-518.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. "Family Welfare Cultures." *The Quarterly Journal of Economics* 129, no. 4 (2014): 1711-1752.

Dalrymple, Theodore. *Society is Broken.* The Centre for Independent Studies, 2016.

Darling-Hammond, Linda. "Keeping Good Teachers: Why It Matters, What Leaders Can Do." *Educational Leadership* 60, no. 8 (2003): 6-13.

de Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers. "Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia." *The Quarterly Journal of Economics* 133, no. 2 (2018): 993-1039.

Dee, Thomas S., and James Wyckoff. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." *Journal of Policy Analysis and Management* 34, no. 2 (2015): 267-297.

Ding, Min. "An Incentive-Aligned Mechanism for Conjoint Analysis." *Journal of Marketing Research* 44, no. 2 (2007): 214-223.

Dolton, Peter, & Van der Klaauw, Wilbert. (1999). "The Turnover of Teachers: A Competing Risks Explanation." *The Review of Economics and Statistics*, *81*(3), 543-550.

Elementary & Secondary Information System. "Table Generator." *National Center for Education Statistics.* https://nces.ed.gov/ccd/elsi (accessed January 28, 2019).

Ehrenberg, Ronald G. "Retirement System Characteristics and Compensating Wage Differentials in the Public Sector." *ILR Review,* 33, no. 4 (1980): 470-483.

Ericson, Keith Marzilli, and David Laibson. "Intertemporal Choice." No. w25358. National Bureau of Economic Research, 2018.

Feng, Li. "Hire Today, Gone Tomorrow: The Determinants of Attrition Among Public School Teachers." Working paper (2005).

Feng, Li, and Tim R. Sass. "Teacher quality and teacher mobility." *Education Finance and Policy* 12, no. 3 (2017): 396-418.

Figlio, David N. "Can Public Schools Buy Better-Qualified Teachers?." *ILR Review* 55, no. 4 (2002): 686-699.

Figlio, David N. "Teacher Salaries and Teacher Quality." *Economics Letters* 55, no. 2 (1997): 267-271.

Fitzpatrick, Maria Donovan. "How Much Are Public School Teachers Willing to Pay for Their Retirement Benefits?" *American Economic Journal: Economic Policy* 7, no. 4 (2015): 165-88.

Freeman, Richard B., and James L. Medoff. "What do unions do." *Indus. & Lab. Rel. Rev.* 38 (1984): 244.

Gilraine, Michael, and Nolan G. Pope. "Making Teaching Last : Long- and Short-Run Value-Added." Working paper (2020).

Glaeser, Edward L., and Giacomo A.M. Ponzetto. "Shrouded Costs of Government: The Political Economy of State and Local Public Pensions." *Journal of Public Economics* 116 (2014): 89-105.

Goda, Gopi Shah, Damon Jones, and Colleen Flaherty Manchester. "Retirement Plan Type and Employee Mobility the Role of Selection." *Journal of Human Resources* 52, no. 3 (2017): 654-679.

Goldhaber, Dan, Roddy Theobald, and Christopher Tien. "Why We Need a Diverse Teacher Workforce." *Phi Delta Kappan,* 100, no. 5 (2019): 25-30.

Goldin, Claudia. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review,* 104, no. 4 (2014): 1091-1119.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. "The Effect of School Resources on Student Achievement." *Review of Educational Research,* 66, no. 3 (1996): 361-396.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis,* 22, no. 1 (2014): 1-30.

Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. "Validating Vignette and Conjoint Survey Experiments against Real-World Behavior." *Proceedings of the National Academy of Sciences,* 112, no. 8 (2015): 2395-2400.

Hainmueller, Jens, and Michael J. Hiscox. "Attitudes Toward Highly Skilled and Low-Skilled Immigration: Evidence from a Survey Experiment." *American Political Science Review,* 104, no. 1 (2010): 61-84.

Hanushek, Eric A. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis,* 19, no. 2 (1997): 141-164.

Hanushek, Eric A. "The Economic Value of Higher Teacher Quality." *Economics of Education Review* 30, no. 3 (2011): 466-479.

Hanushek, Eric A. “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature,* 24, no. 3 (1986): 1141-1177.
Hanushek, Eric A., and Steven G. Rivkin. “Teacher Quality.” *Handbook of the Economics of Education* 2 (2006): 1051-1078.
Harris, Anthony R., Stephen H. Thomas, Gene A. Fisher, and David J. Hirsch. “Murder and Medicine: the Lethality of Criminal Assault 1960-1999." *Homicide Studies* 6, no. 2 (2002): 128-166.
Heckman, James J., John Eric Humphries, and Gregory Veramendi. “Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking.” *Journal of Political Economy* 126, no. S1 (2018): S197-S246.
Hendricks, Matthew D. “Does It Pay to Pay Teachers More? Evidence from Texas.” *Journal of Public Economics,* 109 (2014): 50-63.
Hoxby, Caroline Minter. “How Teachers’ Unions Affect Education Production.” *The Quarterly Journal of Economics* 111, no. 3 (1996): 671-718.
Hoxby, Caroline M. “The Effects of Class Size on Student Achievement: New Evidence from Population Variation.” *The Quarterly Journal of Economics,* 115, no. 4 (2000): 1239-1285.
Imberman, Scott A., and Michael F. Lovenheim. “Incentive Strength and Teacher Productivity: Evidence from a Group-Based Teacher Incentive Pay System.” *Review of Economics and Statistics,* 97, no. 2 (2015): 364-386.
Inman, Robert P. “Public Employee Pensions and the Local Labor Budget.” *Journal of Public Economics,* 19, no. 1 (1982): 49-71.
Jackson, C. Kirabo. “Student Demographics, Teacher sorting, and Teacher Quality: Evidence from the End of School Desegregation.” *Journal of Labor Economics,* 27, no. 2 (2009): 213-256.
Jacob, Brian A., Jonah E. Rockoff, Eric S. Taylor, Benjamin Lindy, and Rachel Rosen. ‘‘Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools.” *Journal of Public Economics,* 166 (2018): 81-97.
Johnston, Andrew C (2024). "Replication data for: *Preferences, Selection, and the Structure of Teacher Pay*." Deposit in Progress. AEA Data and Code Repository. https://www.openicpsr.org/openicpsr/workspace?goToPath=/openicpsr/198086.
Johnston, Andrew C., and Carla Johnston. “Is Compassion a Good Career Move? Nonprofit Earnings Differentials from Job Changes.” *Journal of Human Resources* 56, no. 4 (2021): 1226-1253.
Jones, Michael, and Michael T. Hartney. “Show Who the Money? Teacher Sorting Patterns and Performance Pay across US School Districts.” *Public Administration Review* 77, no. 6 (2017): 919-931.
Kane, Thomas J., and Douglas O. Staiger. *Estimating teacher impacts on student achievement: An experimental evaluation*. No. w14607. National Bureau of Economic Research, 2008.
Koedel, Cory, Diyi Li, Morgan S. Polikoff, Tenice Hardaway, and Stephani L. Wrabel. "Mathematics curriculum effects on student achievement in California." *AERA Open* 3, no. 1 (2017).
Kraft, Matthew A., Eric J. Brunner, Shaun M. Dougherty, and David Schwegman. “Teacher Accountability Reforms and the Supply of New Teachers." *Unpublished manuscript. Retrieved from https://scholar. harvard. edu/files/mkraft/files/kraft_et_al. _2018_teacher_accountability_reforms. pdf* (2018).
Krueger, Alan B. “Experimental Estimates of Education Production Functions.” *The Quarterly Journal of Economics* 114, no. 2 (1999): 497–532.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. “Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis.” *Educational Evaluation and Policy Analysis* 24, no. 1 (2002): 37-62.
Lavy, Victor. “Evaluating the Effect of Teachers’ Group Performance Incentives on Pupil Achievement.” *Journal of Political Economy* 110, no. 6 (2002): 1286-1317.
Lavy, Victor. “Performance Pay and Teachers’ Effort, Productivity, and Grading Ethics.” *American Economic Review* 99, no. 5 (2009): 1979-2011.
Lazear, Edward P. “Educational Production.” *The Quarterly Journal of Economics* 116, no. 3 (2001): 777-803.
Lazear, Edward P. “Performance Pay and Productivity.” *American Economic Review* 90, no. 5 (2000): 1346-1361.
Lazear, Edward P. “Teacher Incentives.” *Swedish Economic Policy Review* 10, no. 2 (2003): 179-214.
Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin. “Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools.” *American Economic Review* 111, no. 7 (2021): 2213-46.
Loeb, Susanna, and Marianne E. Page. “Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor-Market Opportunities and Non-Pecuniary Variation.” *Review of Economics and Statistics* 82, no. 3 (2000): 393-408.
Louviere, Jordan J., David A. Hensher, and Joffre D. Swait. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, 2000.
Lovenheim, Michael F., and Alexander Willen. “The Long-run Effects of Teacher Collective Bargaining." *American Economic Journal: Economic Policy* 11, no. 3 (2019): 292-324.
Maestas, Nicole, Kathleen J. Mullen, David Powell, Till Von Wachter, and Jeffrey B. Wenger. “The Value of Working Conditions in the United States and Implications for the Structure of Wages.” No. w25204. National Bureau of Economic Research, 2018.
Mansfield, Richard K. “Teacher Quality and Student Inequality.” *Journal of Labor Economics* 33, no. 3 (2015): 751-788.
Mas, Alexandre, and Amanda Pallais. “Valuing Alternative Work Arrangements.” *American Economic Review* 107, no. 12 (2017): 3722-59.
McFadden, Daniel. “Econometric Models of Probabilistic Choice.” *Structural Analysis of Discrete Data with Econometric Applications* 198272 (1981).
Mont, Daniel, and Daniel I. Rees. “The Influence of Classroom Characteristics on High School Teacher Turnover.” *Economic Inquiry* 34, no. 1 (1996): 152-167.
Morrissey, Monique. “Teachers and Schools Are Well Served by Teacher Pensions.” *Economic Policy Institute* (2017).
Muralidharan, Karthik, and Venkatesh Sundararaman. “Teacher Performance Pay: Experimental Evidence from India.” *Journal of Political Economy* 119, no. 1 (2011): 39-77.
Nagler, Markus, Marc Piopiunik, and Martin R. West. “Weak Markets, Strong Teachers: Recession at Career Start and Teacher Effectiveness.” *Journal of Labor Economics* (2019).
National Center for Education Statistics. “District Detail For Aldine ISD.” *Search For Public School Districts*. https://nces.ed.gov/ccd/districtsearch/index.asp (accessed January 24, 2019).
Neal, Derek A., and William R. Johnson. “The Role of Premarket Factors in Black-White Wage Differences.” *Journal of Political Economy* 104, no. 5 (1996): 869-895.

Newhouse, Joseph P. “Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital.” *The American Economic Review* 60, no. 1 (1970): 64-74.

Novy-Marx, Robert, and Joshua Rauh. “The Revenue Demands of Public Employee Pension Promises.” *American Economic Journal: Economic Policy* 6, no. 1 (2014): 193-229.

Opper, Isaac M. “Does Helping John Help Sue? Evidence of Spillovers in Education.” *American Economic Review* 109, no. 3 (2019): 1080–1115.

Papay, John P., Eric S. Taylor, John H. Tyler, and Mary E. Laski. “Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data.” *American Economic Journal: Economic Policy* 12, no. 1 (2020): 359-388.

Papay, John P., and Matthew A. Kraft. “Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement.” *Journal of Public Economics* 130 (2015): 105-119.

Petek, Nathan, and Nolan Pope. “The Multidimensional Impact of Teachers on Students.” (2019).

Phipps, Aaron. “Personnel Contracts with Production Uncertainty: Theory and Evidence from Teacher Performance Incentives.” Working paper (2018).

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. “Teachers, Schools, and Academic Achievement.” *Econometrica* 73, no. 2 (2005): 417-458.

Rockoff, Jonah E. “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review* 94, no. 2 (2004): 247-252.

Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger. “Can You Recognize an Effective Teacher When You Recruit One?” *Education Finance and Policy* 6, no. 1 (2011): 43-74.

Rothstein, Jesse. “Does Competition among Public Schools Benefit Students and Taxpayers? Comment.” *American Economic Review* 97, no. 5 (2007): 2026-2037.

Rothstein, Jesse. “Teacher Quality Policy When Supply Matters.” *American Economic Review 105*, no. 1 (2015), 100-130.

Sajjadiani, S., A. Sojourner, J. Kammeyer-Mueller & E. Mykerezi (*forthcoming*). “Using Machine Learning to Translate Pre-Hire Work History into Predictors of Performance and Retention.” *Journal of Applied Psychology*.

Schlechty, Phillip C., and Victor S. Vance. “Do Academically Able Teachers Leave Education? The North Carolina Case.” *The Phi Delta Kappan* 63, no. 2 (1981): 106-112.

Schlechty, Phillip C., and Victor S. Vance. 1982. “The Distribution of Academic Abililty in the Teaching Force: Policy Implications.” Phi Delta Kappan 64.

Small, Kenneth A. "Valuation of Travel Time." *Economics of Transportation* 1, no. 1-2 (2012): 2-14.

Smith, Robert Stewart. “Compensating Differentials for Pensions and Underfunding in the Public Sector.” *The Review of Economics and Statistics* (1981): 463-468.

Staiger, Douglas O., and Jonah E. Rockoff. “Searching for Effective Teachers with Imperfect Information.” *Journal of Economic Perspectives* 24, no. 3 (2010): 97-118.

Stinebrickner, T. R.. “A Dynamic Model of Teacher Labor Supply.” *Journal of Labor Economics*, *19*.1 (2001), 196-230.

Texas Education Agency. “STAAR Aggregate Data for 2017–2018.” *Student Testing and Accountability.* https://tea.texas.gov/Student_Testing_and_Accountability/Testing/State_-of_Texas_Assessments_of_Academic_Readiness_(STAAR)/STAAR_Aggregate_Data_for_2017-2018/ (accessed on January 25, 2019).

Todd, Petra E., and Kenneth I. Wolpin. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." *American Economic Review* 96, no. 5 (2006): 1384-1417.

Train, Kenneth E. *Discrete Choice Methods with Simulation*. Cambridge university press, 2009.

U.S. Bureau of Labor Statistics (2020), "Employer Costs for Employee Compensation Survey," *Economic News Release*, https://www.bls.gov/news.release/ecec.nr0.htm (accessed February 2, 2021).

U.S. Department of Education (2016), "Enrollment, Poverty, and Federal Funds for 120 largest School Districts, by Enrollment Size in 2016," *Digest of Education Statistics*, table 215.30, https://nces.ed.gov/programs/digest/d16/tables/dt16_215.30.asp (accessed January 24, 2019).

Walters, Christopher R. "The Demand for Effective Charter Schools." *Journal of Political Economy* 126, No. 6 (2018): 2179-2223.

Watlington, Eliah, Robert Shockley, Paul Guglielmino, and Rivka Felsher. "The High Cost of Leaving: An Analysis of the Cost of Teacher Turnover." *Journal of Education Finance* (2010): 22-37.

Weaver, W. Timothy. "In Search of Quality: The Need for Talent in Teaching." *The Phi Delta Kappan* 61, no. 1 (1979): 29-46.

Weller, Christian. "Win-Win: Pensions Efficiently Service American Schools and Teachers." *National Institute on Retirement Security* (2017).

Wheelan, Charles. *Naked Economics*. WW Norton & Company, 2019.

Willen, Alexander. "Decentralization of Wage Determination: Evidence from a National Teacher Reform." *Journal of Public Economics* 198 (2021): 104388.

Winters, Marcus A., and Joshua M. Cowen. "Would a Value-added System of Retention Improve the Distribution of Teacher Quality? A Simulation of Alternative Policies." *Journal of Policy Analysis and Management* 32, no. 3 (2013): 634-654.

Wiswall, Matthew. "The Dynamics of Teacher Quality." *Journal of Public Economics*, 100 (2013): 61-78.

Wiswall, Matthew, and Basit Zafar. "Preference for the Workplace, Investment in Human Capital, and Gender." *The Quarterly Journal of Economics* 133, no. 1 (2017): 457-507.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. MIT press, 2010.

FIGURE 1—SAMPLE QUESTION

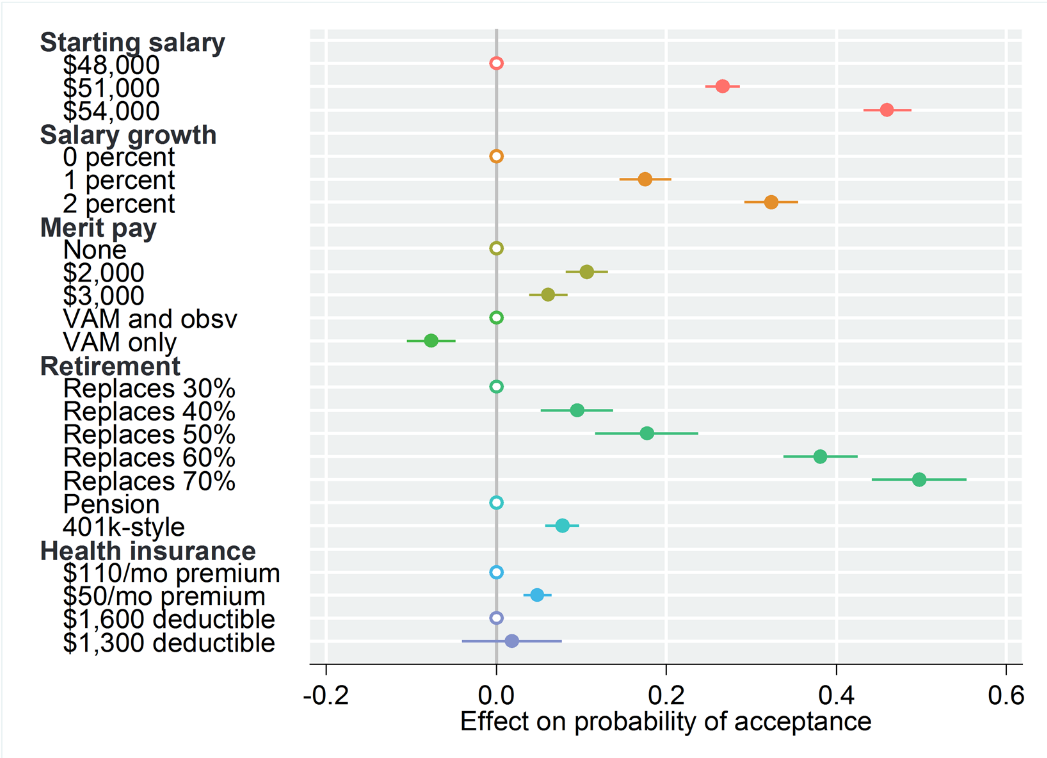
If two schools that were identical in every other way made the following offers, which would you prefer:

	School 1	**School 2**
Starting salary:	$52,850	$46,850
Health plan:	$1,400 deductible; $40 monthly premium	$1,250 deductible; $90 monthly premium
Salary growth:	2.0% each year	4.0% each year
Reward:	Teachers receive $1,000 reward if they are in the top 25% of the school based on principal ratings and student growth	Teachers receive $2,000 reward if they are in the top 25% of the school based on principal ratings and student growth
Retirement:	A pension that replaces 65% of your salary in retirement if you stay 30 years	A pension that replaces 35% of your salary in retirement if you stay 30 years
	○	○

Note: This figure presents a sample question answered by teacher respondents. There were three sets of questions: a set about compensation structure, a set about working conditions, and a set about student and principal attributes. Each set of questions is called a "deck" in the conjoint literature. I generated 30 questions for each deck. Each respondent was randomly assigned four questions from the "compensation" deck, four questions from the "working conditions" deck, and three questions from the "student and principal attribute" deck for a total of eleven questions. In the data, each row represents an option considered in one question by a single respondent, and choice by a respondent is indicated with a binary variable. As seen in this example, questions always pose a tradeoff to the respondent which allows the analyst to efficiently draw out information on the structure of respondent preferences.

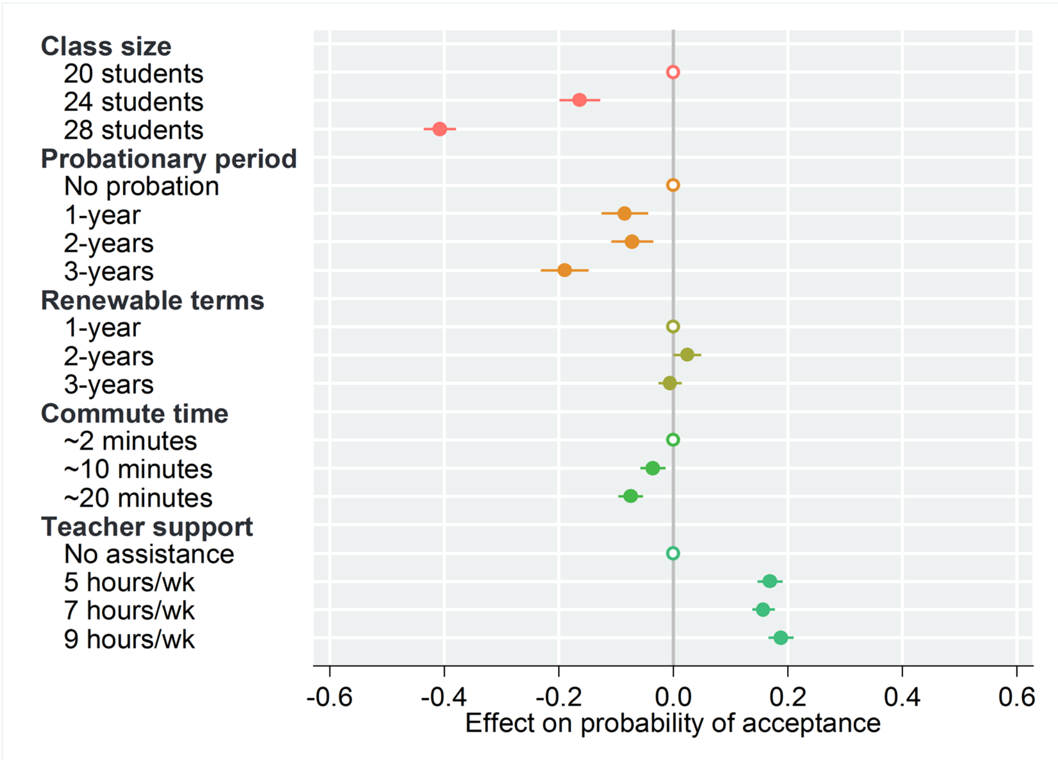
FIGURE 2—EFFECTS OF COMPENSATION ATTRIBUTES

ON THE PROBABILITY THAT TEACHERS SELECTS A JOB OFFER



Note: This figure shows how variation in job-offer attributes affect the probability a respondent chooses a presented option (conditional on all the other attributes), using responses in the "compensation" deck. In the data, each row represents a single job option considered by an individual respondent, so each question a teacher responds to generates two rows in the data, one for each option presented. A respondent's choice of which job offer she prefers is indicated with a binary variable. Intuitively, I regress the choice indicator on the attributes of each job option, and I condition on question/choice-set fixed effects to account for the fact that teachers make selections "within" a question/choice set. Continuous values are segmented into bins for the purpose of this figure. The dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from a least-squares regression. The unfilled dots on the zero line denote the (omitted) reference category for each job-offer attribute. N = 31,820, one row for each respondent-question-option combination.

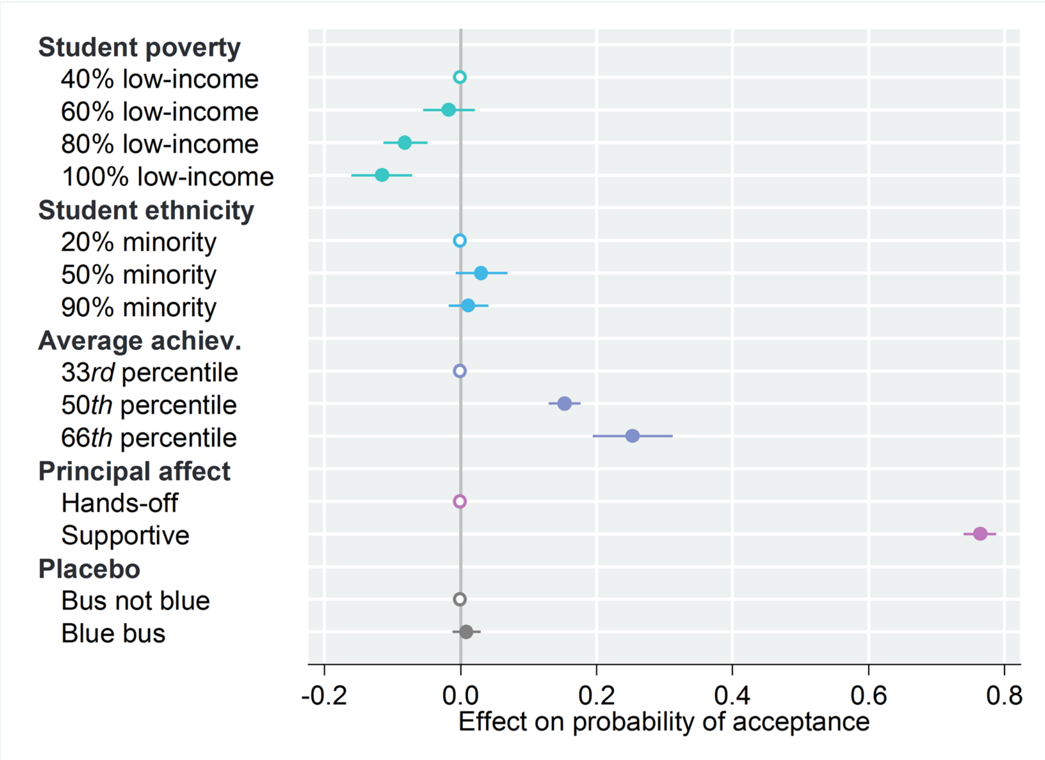
FIGURE 3—EFFECTS OF WORKING-CONDITION ATTRIBUTES ON THE PROBABILITY THAT TEACHERS SELECTS A JOB OFFER



Note: This figure shows how variation in job-offer attributes affect the probability a respondent chooses a presented option (conditional on all the other attributes), using responses in the "working conditions" deck. In the data, each row represents a single job option considered by a respondent, so each question a teacher responds to generates two rows in the data, one for each option presented. A respondent's choice of which job offer she prefers is indicated with a binary variable. Intuitively, I regress the choice indicator on the attributes of each job option, and I condition on question/choice-set fixed effects to account for the fact that teachers make selections "within" a question/choice set. Continuous values are segmented into bins for the purpose of this figure. The dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from a least-squares regression. The unfilled dots on the zero line denote the (omitted) reference category for each job-offer attribute. I have used the estimates on starting salary in this deck to harmonize the magnitudes so that estimates are directly comparable to those in figures 2 and 4. N = 31,574, one row for each respondent-question-option combination.

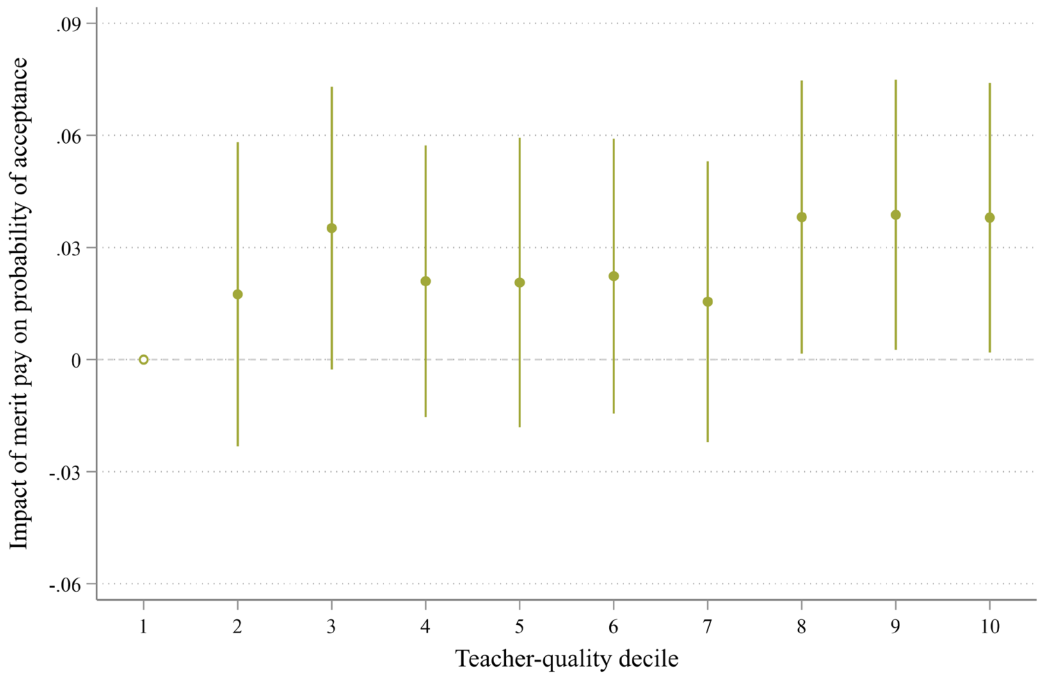
FIGURE 4—EFFECTS OF STUDENT AND PRINCIPAL ATTRIBUTES

ON THE PROBABILITY THAT TEACHERS SELECTS A JOB OFFER



Note: This figure shows how variation in job-offer attributes affect the probability a respondent chooses a presented option (conditional on all the other attributes), using responses in the "students and leadership" deck. In the data, each row represents a single job option considered by a respondent, so each question a teacher responds to generates two rows in the data, one for each option presented. A respondent's choice of which job offer she prefers is indicated with a binary variable. Intuitively, I regress the choice indicator on the attributes of each job option, and I condition on question/choice-set fixed effects to account for the fact that teachers make selections "within" a question/choice set. Continuous values are segmented into bins for the purpose of this figure. The dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from a least-squares regression. The unfilled dots on the zero line denote the (omitted) reference category for each job-offer attribute. I have used the estimates on starting salary in this deck to harmonize the magnitudes so that estimates are directly comparable to those in figures 2 and 3. N = 23,678, one row for each respondent-question-option combination.

FIGURE 5—DIFFERENTIAL EFFECT OF MERIT PAY ON THE PROBABILITY THAT TEACHERS ACCEPT A JOB OFFER



Note: This figure shows how variation in performance pay affects the probability a respondent chooses a presented offer for teachers with different math value-add (VA). I separate respondents into ten deciles where those in the top decile have the highest value-add and those in the bottom decile have the lowest. I implement a regression, like those implemented before, where choice is regressed on offer attributes and I interact performance pay with indicators for VA deciles 2 through 10, making decile 1 the omitted category. The coefficients plotted above represent the differential effect of an additional $1,000 in performance pay on the probability that a teacher will accept a job offer. The figure demonstrates that the top three deciles of teachers have significantly stronger preferences for performance pay than bottom decile teachers. The standard errors are clustered by respondent and the 95% confidence intervals are presented. N = 11,014, one row for each respondent-question-option combination among respondents with a math VA.

TABLE 1—SUMMARY STATISTICS ON OFFER ATTRIBUTES FOR CONJOINT EXPERIMENT

Variable	Deck(s)	Average	Std. Dev.	Units	N
Choice	1, 2, 3	0.50	(0.50)	Indicator	87,078
Starting Salary	1, 2, 3	50.52	(2.84)	$1,000s	87,078
Salary Growth	1	1.44	(0.70)	% growth	31,826
Bonus amount	1	1.25	(1.29)	$1,000s	31,826
VA only	1	0.20	(0.40)	Indicator	31,826
Replacement	1	48.08	(9.30)	% of salary	31,826
401k-style	1	0.37	(0.48)	Indicator	31,826
Premium (yearly)	1	0.78	(0.30)	$1,000s	31,826
Deductible	1	1.48	(0.18)	$1,000s	31,826
Probationary period	2	1.72	(0.93)	Years	31,574
Term length	2	2.26	(0.96)	Years	31,574
Commute time	2	0.19	(0.10)	Hours	31,574
Class size	2	24.55	(3.39)	Students	31,574
Assistance	2	3.25	(3.66)	Hours/week	31,574
Percent low income	3	6.79	(1.86)	10%s	23,678
Percent minority	3	5.62	(2.97)	10%s	23,678
Ave. achievement	3	4.99	(1.65)	10%tiles	23,678
Supportive	3	0.47	(0.50)	Indicator	23,678
Blue bus	3	0.50	(0.50)	Indicator	23,678

Note: This table presents the mean and standard deviation of the experimental data where each observation is a single option considered by a teacher respondent. The units column describes the units of each variable to aid interpretation of regression results. There were 4,262 unique teacher respondents in the choice experiment.

TABLE 2—LINEAR PREFERENCES OVER
COMPENSATION STRUCTURE AND WORKING CONDITIONS

	Choice (Linear Probability)			Choice (Conditional Logit)		
	Coeff	SE	WTP	Coeff	SE	WTP
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: Compensation Deck						
Salary						
Starting salary	0.085**	(0.002)	$1,000	0.402**	(0.007)	$1,000
Salary growth	0.192**	(0.009)	$2,268	0.948**	(0.034)	$2,361
Merit reward						
Bonus amount	0.029**	(0.003)	$346	0.191**	(0.012)	$475
VA only	-0.077**	(0.015)	-$907	-0.218**	(0.051)	-$543
Retirement						
Replacement	0.015**	(0.001)	$173	0.072**	(0.002)	$178
401k-style	0.077**	(0.010)	$907	0.406**	(0.033)	$1,010
Health insurance						
Premium (yearly)	-0.082**	(0.014)	-$971	-0.427**	(0.047)	-$1,063
Deductible	-0.312	(0.212)	-$3,686	-1.010	(0.756)	-$2,514
Panel 2: Working-Conditions Deck						
Contract						
Probationary period	-0.042**	(0.004)	-$501	-0.315**	(0.022)	-$461
Term length	-0.003	(0.004)	-$33	0.011	(0.020)	$16
Working conditions						
Commute time	-0.268**	(0.032)	-$3,173	-2.836**	(0.198)	-$4,159
Class size	-0.050**	(0.001)	-$594	-0.395**	(0.007)	-$579
Assistance	0.022**	(0.001)	$256	0.173**	(0.005)	$253
Panel 3: Students-&-Leaders Deck						
Students						
Percent low income	-0.027**	(0.003)	-$325	-0.116**	(0.010)	-$282
Percent minority	0.003	(0.002)	$40	0.008	(0.006)	$19
Ave. achievement	0.046**	(0.003)	$546	0.236**	(0.011)	$575
Principal affect						
Supportive	0.733**	(0.012)	$8,672	3.068**	(0.042)	$7,466
Placebo						
Blue bus	0.009	(0.010)	$101	0.021	(0.038)	$52

Notes:

Note: * $p < 0.05$, ** $p < 0.001$. This table shows how variation in job-offer attributes affect the probability a respondent chooses a presented option (conditional on all the other attributes), using responses from all three decks. In the data, each row represents a single job option considered by an individual respondent, so each question a teacher responds to generates two rows in the data, one for each option presented. A respondent's choice of which job offer she prefers is indicated with a binary variable. Intuitively, I regress the choice indicator on the attributes of each job option, and I condition on question/choice-set fixed effects to account for the fact that teachers make selections "within" a question/choice set. In addition to presenting the ordinary least squares results in columns (1) through (3), I present the conditional logit estimates in columns (4) through (6). Columns (1) and (4) present the parts-worth utility coefficient, which is the effect of job-offer variation on the probability of accepting a presented offer. Columns (2) and (5) present the standard errors of those estimates. And columns (3) and (6) present the willingness to pay for a unit of each attribute, which is found by dividing the parts-worth utility of a given attribute by the parts-worth utility of salary in the same deck, and then multiplying by $1,000. Standard errors are clustered by respondent to account for correlated responses by the same teacher. I have used the estimates on starting salary in decks 2 and 3 to harmonize the magnitudes so that estimates are directly comparable across decks. N = 31,574, one row for each respondent-question-option combination. Regression summaries: Deck 1: N=31,820, %Predicted=64, R-squared=0.19; Deck 2: N= 31,574, %Predicted=64, R-squared=0.28; Deck 3: N=23,678, %Predicted=62, R-squared=0.36.

TABLE 3—DO PRINCIPALS MITIGATE DIFFICULT WORK SETTINGS?

	Choice		
	(1)	(2)	(3)
Supportive Principal (SP)	0.575**	0.794**	0.683**
	(0.009)	(0.054)	(0.067)
Achievement pctl.	0.036**	0.058**	0.067**
	(0.003)	(0.006)	(0.006)
Achievement pctl. × SP	.	-0.045**	-0.061**
	.	(0.011)	(0.012)
Poverty rate	-0.022**	-0.020**	-0.033**
	(0.002)	(0.003)	(0.005)
Poverty rate × SP	.	.	0.027*
	.	.	(0.009)
Observations	*23,678*	*23,678*	*23,678*
R-squared	*0.365*	*0.366*	*0.366*

Note: * p < 0.05, ** p < 0.001. This table shows how variation in student attributes and principal support affect the probability of choosing a presented offer (conditional on all the other attributes), using responses from the "students and principal attribute" deck. In the data, each row represents a single job option considered by a respondent, so each question a teacher responds to generates two rows in the data, one for each option presented. A respondent's choice of which job offer she prefers is indicated with a binary variable. Intuitively, I regress the choice indicator on the attributes of each job option, and I condition on question/choice-set fixed effects to account for the fact that teachers make selections "within" a question/choice set. "Supportive Principal" is a binary variable indicating whether principals are "supportive with disruptive students." (The alternative is that the principal is "hands off with disruptive students"). "Achievement pctl." is the listed average achievement percentile of the school's students in the offer, and "Poverty rate" is the low-income share of the school's students in the offer. Standard errors are clustered by respondent to account for correlated responses by the same teacher across questions. The interaction terms show that a supportive principal essentially erases teacher aversion to low-performing and high-poverty schools.

TABLE 4—TEACHER PREFERENCES BY QUALITY

	Choice (1)	Choice (2)	Choice (3)
Salary	0.091**	0.092**	0.086**
	(0.004)	(0.004)	(0.003)
Salary x quality	0.011	0.010	-0.002*
	(0.009)	(0.016)	(0.001)
Growth	0.183**	0.183**	0.185**
	(0.012)	(0.012)	(0.009)
Growth x quality	0.018	0.047	-0.001
	(0.023)	(0.036)	(0.003)
Merit pay	0.046**	0.047**	0.027**
	(0.006)	(0.006)	(0.004)
Merit pay x quality	**0.040****	**0.055***	**0.005****
	(0.015)	**(0.025)**	**(0.002)**
VA only	-0.071**	-0.072**	-0.075**
	(0.025)	(0.025)	(0.018)
VA only x quality	-0.025	-0.025	-0.001
	(0.038)	(0.059)	(0.004)
Replacement	0.014**	0.014**	0.014**
	(0.001)	(0.001)	(0.001)
Replace. x quality	0.001	-0.002	0.000*
	(0.002)	(0.003)	(0.000)
401k style	0.061**	0.062**	0.076**
	(0.018)	(0.018)	(0.013)
401k x quality	-0.005	-0.006	0.005
	(0.041)	(0.067)	(0.004)
Premium (yearly)	-0.050*	-0.049*	-0.090**
	(0.024)	(0.024)	(0.018)
Premium x quality	0.083	0.135	-0.002
	(0.072)	(0.118)	(0.008)
Deductible	-0.508**	-0.507**	-0.549**
	(0.058)	(0.058)	(0.042)
Deductible x quality	-0.038	-0.306	0.001
	(0.098)	(0.163)	(0.010)
Quality measure	Math VA	Read VA	Danielson

Question FE	X	X	X
R-squared	0.207	0.206	0.197
N	10,398	10,398	18,838

Note: * $p < 0.05$, ** $p < 0.01$. This table shows whether preferences for compensation structure and working conditions differ by measures of teacher quality, using responses from all three decks. In the data, each row represents a single job option considered by a respondent, so each question a teacher responds to generates two rows in the data, one for each option presented her. A respondent's choice of which job offer she prefers is indicated with a binary variable. Intuitively, I regress the choice indicator on the attributes of each job option, and I condition on question/choice-set fixed effects to account for the fact that teachers make selections "within" a question/choice set. The main effect coefficients reflect how a unit change in an attribute affects respondent choice when the quality measure is zero. The interaction terms show how preferences for each attribute changes as teacher quality changes. I show three quality measures in this table: math value add, reading value add, and a Danielson score which is generated from principal observations. I find that teachers with high VA have broadly similar preferences to those that have low VA, except in one regard. High VA teachers have stronger preference for options including performance pay. Importantly, teachers in Aldine do not receive VA estimates or reports, so special preference for performance pay among high VA teachers probably reflects private information. Standard errors are clustered by respondent to account for correlated responses by the same teacher across questions. Danielson scores tend to be weakly correlated with VA, having an R-squared of 0.026–0.049 when predicting math VA (depending on the controls when calculating VA) and an R-squared of 0.008–0.011 when predicting reading VA.

TABLE 5—SIMULATED COMPENSATION STRUCTURE UNDER VARIOUS OBJECTIVES

	Status quo	**Teacher-utility optimal**	**Teacher-retention optimal**	**Student-achievement optimal**
	(1)	**(2)**	**(3)**	**(4)**
Starting salary	$50,000	$75,655	$66,688	$66,771
Salary growth	1.8%	0.0%	1.3%	1.3%
Merit pay	$0	$1,477	$1,487	$5,000
VA-only merit	0	0	0	1
Replacement rate	69.0%	55.5%	56.6%	56.9%
Defined contribution	0	1	1	1
Insurance subsidy	$3,960	$0	$0	$0
Class size	28.7	30.0	30.0	30.0
Teacher utility	$79,200	**$96,300**	$90,800	$85,000
Teacher experience	9.03 years	10.9 years	**11.0 years**	10.0 years
Student achievement	0.092σ	0.135σ	0.135σ	**0.183σ**

Note: In this table, I compare the status quo compensation bundle (column 1) to the bundle that maximizes teacher welfare (column 2), the bundle that maximizes teacher retention (column 3), and the bundle that maximizes student achievement (bundle 4). I maximize each objective subject to the current budget constraint so each bundle costs the same. The utility objective function is simply the estimated parts-worth utility function estimated from the choice experiment with quadratic terms in salary growth, merit pay, the replacement rate, and class size. I divide calculated utility by the coefficient on salary (and multiply by $1,000) to present utility in terms of equivalent annual cash transfer (for instance, $79,200 in teacher utility means that the annual bundle is valued by teachers equal to a $79,200 pure-cash transfer with no fringe benefits). The retention objective function simulates the retention profile over the life course using Hendricks (2014) estimates of retention of retention response to utility improvements over the teacher life cycle and calculates the average simulated experience of teachers. The achievement objective function is a function of the simulated experience profile, the class size, and the effects of merit pay on selection and effort of a given bundle. The three rows at the bottom of the table show the calculated teacher utility, average teacher experience, and student achievement gains produced by each bundle. These are simply the objective function calculations applied to each listed bundle.