I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14831

Preferences, Selection, and the Structure of Teacher Pay

Andrew C. Johnston

NOVEMBER 2021

DISCUSSION PAPER SERIES

IZA DP No. 14831

Preferences, Selection, and the Structure of Teacher Pay

Andrew C. Johnston
University of California, Merced and IZA

NOVEMBER 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Preferences, Selection, and the Structure of Teacher Pay[1]

I conduct a discrete-choice experiment with responses linked to administrative teacher and student records to examine teacher preferences for compensation structure and working conditions. I calculate willingness-to-pay for a rich set of work attributes. High-performing teachers have similar preferences to other teachers, but they have stronger preferences for performance pay. Taking the preference estimates at face value I explore how schools should structure compensation to meet various objectives. Under each objective, schools appear to underpay in salary and performance pay while overpaying in retirement. Restructuring compensation can increase both teacher welfare and student achievement.

JEL Classification: I20, J32, J45, M50

Keywords: teacher labor markets, compensation structure, teacher quality

Corresponding author:
Andrew C. Johnston
University of California, Merced
Department of Economics
5200 Lake Road
Classroom Office Building
California
USA

E-mail: acjohnston@ucmerced.edu

1 I am grateful to David Card, Damon Clark, Gordon Dahl, Michael Dinerstein, Mark Duggan, Laura Giuliano, Alex Mas, Jesse Rothstein, Kenneth Shores, Isaac Sorkin, and Christopher Walters, 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. Views expressed are those of the author and should not be attributed to The University of California.

I. Introduction

Human capital is a pervasive factor shaping income, inequality, and growth (Neal and Johnson 1996; Barro 2001; Chetty, Friedman, and Rockoff 2014b). In the formation of human capital, teachers are quite possibly the most formative public input (Darling-Hammond 2003; Rockoff 2004; Rivkin, Hanushek, and Kain 2005). High-performing teachers promote greater achievement and non-cognitive skills, which translate to higher earnings and better outcomes in adulthood (Chetty, Friedman, and Rockoff 2014; Opper 2019; Gilraine and Pope 2020). The effect teachers have on human capital varies widely and improves with experience. This implies the focal importance of teacher *selection* and *retention* in building the stock of human capital.

In the United States, governments spend nearly $1 trillion per year on K–12 education, the principal cost of which is personnel (80 percent). Teachers are subject to a distinctive compensation structure: low salary, generous retirements, and no performance incentives. Because public schools operate with significant market power and without a market test, it is not obvious that schools will structure compensation optimally (Hoxby 2000; Rothstein 2007; de Ree et al. 2018). Political incentives may distort compensation away from a natural optimum (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 a rich set of attributes of compensation and working conditions using a discrete-choice experiment. Schools can improve the appeal of teaching by shifting resources toward features teachers value highly relative to cost. Second, I measure whether high-performing teachers have distinct preferences for pay structure and working conditions. If so, policymakers can improve selection. Third, I use these estimates to calculate counterfactual compensation policies that maximize policy objectives, subject to the current budget constraint. The model allows us to identify opportunities for efficiency gains.

Normally, economists measure preferences by collecting data on the menu of options available when agents choose and compare choice across options with different characteristics (Train 2009; Wiswall and Zafar 2017). This approach is unfruitful among teachers: (1) teachers almost never entertain simultaneous offers;[2] (2) observed school attributes are endogenous to

[2] Teachers rarely entertain simultaneous offers because offers are made over most of the year and explode on the day they are extended.

important unobserved ones; and (3), perhaps most critically, variation in compensation structure among schools is extremely limited. Contracts are essentially uniform across schools, and many important features of the policy space do not vary.[3]

To address these challenges, I use a discrete-choice experiment in which primary and secondary teachers select hypothetical offers varying compensation structure, contract type, and working conditions. Teachers have a consequential reason to reveal their preferences: teachers knew the survey was delivered to inform the district's new pay regime, and their responses affected the policy (Carson et al. 2000). The response rate was high (97 percent). Varied attributes included starting salary, salary growth, retirement generosity, defined-benefit or defined-contribution plans, health insurance premium and deductible, performance pay, the basis of evaluation, time-to-tenure, class sizes, administrative support, student demographics, and principal support. Teacher choices illuminate the structure of their preferences. By estimating preferences over several facets of the work setting, I can calculate the "optimal" compensation structure, something not possible by estimating preferences one element at a time in different settings.

Research has validated this type of experiment in a number of labor settings—hypothetical choices match real-world ones (Mas and Pallais 2017; Wiswall and Zafar 2018; Maestas et al. 2018; Hainmueller et al. 2014). In line with these findings, my experiment elicits responses that match each available benchmark. 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. The discount rate that rationalizes teachers' salary-retirement tradeoff is the same as that estimated in the literature on discounting (Best et al. 2018; Ericson and Laibson 2018). And the cost of commuting implied by teacher choices matches a developed urban literature estimating the same (Small 2012; Mas and Pallais 2017). Success on these benchmarks instills confidence in more novel estimates.

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 equal to a $6,000 increase in salary.[4] Teachers prefer riskier (though portable) defined-contribution retirement plans over a traditional

[3] State policy and common union influence generate similar compensation structures across districts. Within district, compensation is totally uniform.

[4] Here, base starting pay is $50,000 for a new teacher with a bachelor's degree.

pension, holding expected replacement constant. An additional year of probationary status before tenure is equivalent to a salary reduction of 1 percent. These findings are new, and I provide novel estimates on the WTP for a broad array of other school attributes. The attribute teachers most value is a "supportive" principal, valued equal to a 17-percent increase in salary. A supportive principal erases 90 percent of the disutility of teaching in a low-achieving school and reduces the cost of teaching in a low-income setting by 85 percent, suggesting attentive managers reduce the perceived cost of these settings.

Policymakers can improve the appeal of teaching by shifting resources into compensation and amenities that teachers prefer relative to cost. Teachers value defined-benefit retirement income at 60 percent of its cost, and they value defined-contributions retirement income at 104 percent of its cost.[5] The cost of reducing class size is seven times greater than teachers' WTP for smaller classes. For health-insurance subsidies, the WTP is equal to the cost of provision. And teachers on average value a unit of performance pay at a third more than its cost, which doesn't appear to be the result of optimism.

Forecasting which teachers will be most effective is challenging (Hanushek 1986, 1997; Greenwald et al. 1996; Rockoff et al. 2011; Jacob et al. 2018; Sajjadiani et al. 2019). If high-performing teachers have distinctive preferences for conditions controlled by policy, policymakers can affect selection by structuring the work setting to conform to the preferences of high performers (Ballou 1996; Hanushek 2011).[6] Using administrative student-achievement data to calculate value-added measures for teachers (VA) and administrative records on principal evaluations, I find that high-performing teachers have broadly similar preferences to their colleagues except in one regard. Excellent teachers prefer jobs that include the opportunity to earn performance pay. Top-decile teachers value performance pay four times more than lowest-decile teachers. High-performers are 22 percent more likely than low-performers to select an offer providing $3,000 in performance pay which induces favorable selection in retention.[7]

Taking the preference estimates at face value, I examine how schools should structure pay to achieve various objectives. Schools may seek to maximize (i) teacher welfare, (ii) teacher

[5] An additional ten-point replacement rate in pension equivalent to a $1,730 salary increase, less than its cost of $2,870 per year (see Fitzpatrick, 2015). The cost of the same retirement income in a defined-contributions plan is $1,663.

[6] Over time, the effect may be especially pronounced since the preferred compensation differentially retains high-performing teachers who may also prefer work settings inhabited by other high-caliber colleagues (Feng and Sass 2017). 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).

[7] It is less clear whether merit pay would affect sorting *into* the profession if prospective teachers don't have private information on their teaching ability.

retention, or (iii) student achievement. I estimate teacher utility, map utility to attrition decisions (Hendricks 2014), and calibrate a model of schools' achievement production function (Krueger 1999; Papay and Kraft 2015; Imberman and Lovenheim 2015). Whether maximizing teacher utility, teacher retention, or student achievement, the results suggest teachers are underpaid in salary and performance pay while being overpaid in retirement benefits. Restructuring what teachers are paid (subject to the current budget constraint) to maximize their utility generates a 20 percent increase in teacher welfare, the equivalent of a permanent $17,000 raise. Improving teacher welfare increases teacher retention substantially, and student learning rises by 0.04 student standard deviations (σ) per year. Structuring pay to maximize teacher retention introduces higher salary growth to help retain teachers who already have a stock of experience. This bundle raises the average teacher experience by 22 percent in equilibrium and increases the odds of a student having a veteran teacher by 33 percent.

The pay structure that maximizes achievement is similar to the structure that maximizes retention, but it exposes teachers to more performance pay. This structure differentially retains high-performing teachers such that students are 24 percent more likely to have a teacher from the top decile of the original performance distribution than they are in the status quo. The achievement-optimal structure improves student achievement by 0.09 student standard deviations per year, though the full effect takes form over decades. Achievement gains are driven largely by performance pay with 10 percent coming from improved overall retention. The pay structure offers a Pareto improvement across agent types: in addition to increasing student achievement, it improves teacher welfare by 7 percent over the status quo. Under each objective, wage increases come at the expense of lower replacement rates in retirement and shifts toward defined-contributions plans, which are less expensive but preferred by teachers in this setting. To gauge how the effect would scale in general equilibrium, I use a natural experiment to identify the share of gains that are from teacher reallocation. This exercise suggests significant local market power by districts and implies the general-equilibrium effect on achievement would be 10 percent smaller than the partial-equilibrium one.

The model necessarily makes strong assumptions. These assumptions are not unreasonable, and I find support for them in the data. For instance, the model assumes marginal teachers have similar preferences to inframarginal ones. If not, the simulations will be inaccurate for two reasons. First, if a counterfactual policy changes the selection (and preferences) of entrants, the solution is not stable. Second, if potential leavers have dissimilar preferences to inframarginal teachers, retention will be modeled badly. To assess, I (1) test whether teachers who eventually leave the district have the same preferences as those who remain; and (2) survey college students near the district and test whether preferences differ between students who plan to become teachers and those on the margin. In each case, preferences among marginal and inframarginal teachers are indistinguishable. These findings support the view that the counterfactuals are robust to changing selection on entry and the utility estimates apply to marginal teachers (the relevant group) for retention calculations. The homogeneity of preferences implies that teachers have similar preferences for compensation structure as other workers, they only differ in their taste for teaching.

This study contributes to several literatures exploring teacher compensation, teacher quality, and selection.[8] Boyd et al. (2013) use a model to estimate teacher preferences for school characteristics from equilibrium matches. Rothstein (2015) calibrates a model of the teacher labor market and shows that performance incentives or greater dismissals expose teachers to risk such that the practical costs of those policies exceed the likely benefits. Hendricks (2014) exploits natural variation in compensation to estimate the retentive effects of compensation. Fitzpatrick (2015) and Biasi (2020) exploit reforms to measure how teachers value the marginal dollar of their pensions. Biasi (2021) shows that a state reform to individuate teacher pay leads to improved selection and effort among teachers. Finally, Brown and Andrabi (2021) find that performance pay induces positive selection into schools in an experimental design in Pakistan.

In this study, I estimate preferences using a choice experiment and those preferences animate a model of teacher choice grounded in real data. Prior studies are not able to estimate willingness-to-pay for many important components of teacher compensation and working conditions since they do not vary independently in the real world. Where natural-experimental variation allows the analyst to estimate marginal preferences relevant to a given reform, they do not elucidate the whole preference structure, which is needed for counterfactual analysis. The key contribution of this study is to circumvent these issues by creating a transparent choice environment to measure teacher preferences over a broad set of important elements of the work setting, including dimensions for which there is insufficient variation in the natural world. By measuring preferences for a comprehensive set of attributes in a broad policy space, I calculate counterfactual structures and evaluate their effect on teacher welfare, teacher retention, and student achievement. This is the first

[8] The papers in these traditions are extensive. To name only several: Antos and Rosen 1975; Boyd et al. 2013; Biasi 2019; Hanushek 1986; Ballou and Podgursky 1995, 1997; Figlio 1997; Loeb and Page 2000; Hendricks 2014; Rockoff 2004; Hanushek and Rivkin 2006; Jackson 2009; Chetty, Friedman, and Rockoff 2014; Lazear 2000; Staiger and Rockoff 2010; Winters and Cowen 2013; Rothstein 2015; Baron 2020; Brown and Andrabi 2020.

study to use a discrete-choice experiment to calculate policy experiments for compensation structure and working conditions, and it does so in a profession of pivotal importance.

II. Experimental Design and Econometric Framework

The Empirical Challenge

When economists set out to estimate preferences, they collect data on the choices people make and the options available to them at the time of choosing. Unfortunately, the records needed to construct menus from which teachers select offers are unavailable. Districts have no reason to keep records of offers made, and—because of the structure of the market—teachers tend not to receive competing offers.[9] If these records were collected, omitted variables would make it impossible to isolate the causal effect of each attribute. As an example, salary would appear to be more preferred than it really is if schools that pay more also had better amenities. Alternatively, salary would appear less preferred than it really is if schools pay more to compensate for unobservably difficult work settings (Antos and Rosen 1975). In either case, the resulting estimates would not be useful for predicting the effect of policy experiments.

Even if these challenges were somehow surmountable, the results would not be particularly informative. There is essentially no independent variation in most of the school attributes that form a work setting for teachers. It is common for competing schools to have identical compensation structures, tenure timelines, and rules governing working conditions like class size. Even across districts, variation is extremely limited by market concentration, statewide policy, and the common effect of union bargaining. Districts within a state share a pension program. Where variation sometimes exists at the borders between districts, the wealthier district usually offers a work setting that exceeds the neighboring district in every dimension, providing no information on preferences other than what was already known: that more is usually preferred.[10]

How, then, to study teacher preferences? I use a discrete-choice experiment in the field. I generate hypothetical job offers that randomly vary compensation structure and working

[9] The job market is highly decentralized, so schools make offers at widely varying times; since offers explode within 24 hours, teachers rarely entertain multiple concurrent offers. If these records could be assembled, the resulting estimation would reflect the preferences of a relatively distinct subsample of highly sought-after teachers. In the dozens of districts interviewed, none kept records of offers made, precluding the assembly of what offers from which a teacher selected. One alternative is to work though software companies providing application and hiring software to multiple school districts, called consortiums. These software systems include the functionality to extend and accept offers through their interface, but less than one percent of offers were delivered through the software, and many appear to have been in error. Essentially no one accepted their offer through the interface.

[10] This empirical problem is inherent to the setting: wealthy areas often create their own district so as not to subsidize poorer areas. For instance, the wealthy parts of Los Angeles—Beverly Hills, Manhattan Beach, Santa Monica—each have their own district visibly gerrymandered out of the largely poor Los Angeles Unified School District.

conditions and measure teacher choice. An example of the questions asked is presented as figure 1. The experiment neatly addresses the empirical challenges endemic to the question. First, the setting allows us to directly observe menus so that we can see the options from which teachers select. Second, the experiment addresses omitted variables using a controlled experimental setting in which there are no factors unobserved. And third, the environment allows me to introduce independent variation in important policy variables that don't exist or vary in the natural world. These are precisely the issues that make the study of teacher preferences challenging (and in some cases impossible) with naturally occurring records.

Evaluating the Validity of Discrete-Choice Experiments

The discrete-choice experiment, sometimes called a conjoint, is a tool initially developed to measure consumer preferences and forecast demand for components of a prospective product or service. The design started in marketing and is valued because these experiments faithfully predict real-world purchasing behavior and broader market shares (Beggs, Cardell, and Hausman 1981; Green and Srinivasan 1990; Hainmueller, Hopkins, and Yamamoto 2013). Since then, a rich literature has been developed in public, environmental, and experimental economics to assess under what circumstances subjects reveal their preferences truthfully. Based on both theory and evidence, there is good reason to believe responses reflect truth-telling in my setting.

A variety of features of my experiment conduce truthful responses in hypothetical choice. (1) Preference estimates from hypothetical choices where *tradeoffs* are emphasized align with preference estimates from incentive-compatible mechanisms. (2) Recent studies in labor find the career preferences elicited in incentivized settings match those elicited in hypothetical ones. (3) Hypothetical choices where people have experience with the context produce reliable responses. (4) The actual preferences elicited in my experiment closely match the theoretical and empirical benchmarks available. (5) Social-desirability bias is credibly avoided. And, (6) in this setting, there is *consequence* to teachers' choices because the district using the results of the survey to reform its compensation. Therefore, each question is a kind of referendum, the response to which is incentive compatible under a few weak conditions. I expand on the main points.

First, whereas questions asking for open-ended willingness-to-pay (e.g., *how much would you pay for…?*) introduces hypothetical bias, choices that make tradeoffs salient appear to produce the same results as truth-telling mechanisms. For instance, hypothetical auctions produce higher valuations than truth-telling Vickery auctions, but a hypothetical auction that merely emphasizes tradeoffs (asking subjects to visualize paying one's stated valuation) produces the same valuation as the Vickery auction (List 2001). In the same arc, hypothetical choices that emphasize tradeoffs produce indistinguishable estimates from incentive-compatible referenda for public goods, eliminating hypothetical bias (Cummings and Taylor 1999). In discrete choice experiments, too—where tradeoffs are explicitly presented—subjects do not appear to misrepresent their preferences (Vossler, Doyon, and Rondeau 2012). In my discrete choice experiment—where each choice presents tradeoffs—it's therefore likely that teachers provide their preferences truthfully.

Second, recent experiments fielded in labor and public find that the same preferences are found when choice is incentivized or purely hypothetical. Mas and Pallais (2017) present a menu of job alternatives in a real labor market and find that the revealed preferences there are exactly those implied by answers to hypothetical questions. Wiswall and Zafar (2018) find that hypothetical career choices in a lab successfully predict student's eventual career selection two years later. Maestas et al. (2018) find that preferences estimated from hypothetical job choices match the endogenous sorting of workers to jobs. The strongest test of the external validity of conjoint experiments is found in Hainmueller, Hangartner, and Yamamoto (2015). In Switzerland, local citizens vote on whether to naturalize individual migrants using migrant-specific referenda. For each immigrant, citizens cast a secret vote whether to grant permanent status, and citizens are provided detailed demographic information on each candidate migrant: age, sex, origin, language, and integration status. Hainmueller and coauthors compare the results of these real-world referenda to those implied by hypothetical choice. They conclude, "the effects of the applicant attributes estimated from the survey experiments perform *remarkably well* in recovering the effects of the same attributes in the behavioral benchmark [(the referenda)]" (emphasis added). These recent papers provide reason for confidence that discrete choice experiments elicit true preferences, even without incentives.

Third, incentive compatibility seems to matter only when discovering one's preferences requires significant effort, or if subjects have a distinct reason to dissemble;[11] estimates from hypothetical choices align with those from incentivized elicitations in settings where respondents already know their preferences (Camerer and Hogarth 1999; Mas and Pallais 2017; Maestas et al. 2018). Because compensation and working conditions affect a teacher's daily life, they have likely considered their preferences, suggesting no need for new effort to discover them. Several papers

[11] Camerer and Hogarth (1999) remark "In many tasks incentives do not matter, presumably because there is sufficient intrinsic motivation…or additional effort does not matter."

document that experimental valuations approach a neo-classical ideal as subjects gain experience in the setting (List 2003, 2004a, 2004b).

Fourth, I evaluate whether the estimated preferences match various benchmarks. In each benchmark available, the survey performs remarkably well as summarized in the last section and expanded upon in section III.

Fifth, the method avoids the influence of social-desirability bias. There is a large literature documenting that respondents (significantly) alter their answers to present socially desirable responses (Atkin and Chaffee 1972; Campbell 1981; Cotter, Cohen, and Coulter 1982; Finkel, Guterbock, and Borg 1991; Fisher 1993; Krosnick 1999). Surveys where an interviewer is not present conduce truth-telling (Legget et al. 2003; List, Berrens, Bohara, and Kerkvliet 2004; Alpizar et al. 2008). The online survey avoids these issues by providing the subject essentially anonymous privacy. Moreover, the survey design allows the subject to be honest by shrouding sensitive preferences. Subjects are presented a "long list" of attributes, and so they have multiple plausible justifications for any choice made in the conjoint setting (Karlan and Zinman 2012; Hainmueller, Hopkins, and Yamamoto 2014). If a teacher selects an offer with fewer minority students, for instance, she can point to any of the other attributes of the option as her rationale. Respondents enjoy privacy even from the researcher. The analyst cannot infer the preferences of any individual because each respondent makes fewer choices than there are factors (Lowes et al. 2017). Teacher responses are kept confidential and have been reliably private in previous surveys implemented by the consulting group with whom I partnered; thus, teachers have no reason to believe their employer will ever be able to review their individual responses.

Last, there is an actual consequence of teachers' response to the survey, which provides incentives for teachers to respond truthfully. Because each question provided to teachers is essentially a referendum, the dominant strategy is to report one's preferences in earnest (Carson, Groves, and Machina 2000; see also Vossler, Doyon, and Rondeau 2012). Carson and coauthors demonstrate that, for any binary choice where the outcome function is weakly responsive in each agent's message, the dominant strategy is for every agent to report truthfully, selecting the hypothetical offer *if and only if* they prefer that alternative. Several authors show empirically that responses are equivalent, even as they vary the degree of perceived consequentiality to subjects (Bulte et al. 2005, Carson et al. 2006, Herriges et al. 2010).

Implementation

This paper estimates teacher utility over prospective compensation structures, contract terms, and working conditions. 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, which 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.[12] 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.

I consider fourteen attributes recommended by the literature. 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. In this paper, I focus on the estimates for compensation and costly working conditions to examine the effect of compensation structure. I report on results for a few other relevant attributes, and those that allow us to assess the realism of responses. Attributes take on several values, shown in online Appendix table 1.[13]

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 six

[12] 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.

[13] 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.

in a single choice (Green and Srinivasan 1990)), 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 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 2–4.

One 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 a normal 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 stimulates teachers to exhibit preferences for things that have no impact whatever on their welfare. Reliably, I find that teachers express no preference over this irrelevant detail, aiding a preferential interpretation. Uninstructed, subjects may fill in the state space, inferring other characteristics that influence their choice other than those features explicitly described. I frame each question by asking teachers to imagine that two hypothetical job offers are identical in every other way, indicating that the presented school qualities do not relate to unobserved aspects, 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 themselves generates classical measurement error in utility (Wooldridge 2010).[14] Second, the survey is administered digitally, and the option to advance to the next question does not appear for the first few 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 teacher takes to respond to each question; teachers appear to take enough time to read and understand the options, 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, and the estimates are identical in the two subsamples, suggesting that more attention is not associated with different estimates. This alleviates the concern that inattention affects the result.[15]

I deployed the experiment in a large, urban 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, reminder emails, and a lottery for gift cards.

Conceptual and Econometric Framework

Teachers are presented a series of eleven questions in which they choose between two competing job offers. I use their 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. One option may provide a higher salary but comes at the cost of a larger class; a more generous retirement plan accompanies a smaller potential for performance pay. Under weak conditions, the hypothetical job selection data identify job preferences over several factors while standard realized choice 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 vector describing the working conditions, including contract features like the time-to-tenure. For simplicity, I assume utility is additively separable in attributes.

Offers are indexed by j, and there is a finite set of them $j = 1, \ldots$,J. Each offer is characterized by a vector of K attributes: $X_j = [X_{j1}, \ldots, X_{jK}]$. To explore the influence of each factor, I use conditional logistic models as well as linear-probability models to estimate utility, regressing

[14] I confirm this fact in Monte Carlo simulations in both logistic and OLS (not presented).

[15] 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 ($p < 0.001$).

respondent choices on a vector of characteristics while conditioning on choice-set fixed effects to account for the options available to the teacher in each selection:

$$u_{i(X)} = X'_{jS}\beta + \alpha_s + \varepsilon_i \quad (1)$$

Here, teacher i selects option j from choice set S. Parts-worth utilities are denoted β and characteristics of alternative j are given by X_j. 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. In the main analysis, the linear-probability model is marginally better in explaining choice variation and predicting choice. The standard errors are clustered by teacher to account for persistence in preferences across questions by a single respondent. 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

To set the stage for the analysis, I briefly describe the district. 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 take a survey by a consulting firm which, in 2016, included my experiment. The average teacher in the district has nine years of experience, and 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).

III. Teacher Preferences for Compensation and Other Factors

The first step of the paper estimates teacher preferences for compensation and working conditions. By estimating willingness-to-pay, we can later assess whether teacher welfare can be improved by reallocating compensation.

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 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 of each deck.[16] Columns (3) and (6) represent a money metric, calculating how much teachers value a unit of that feature in terms of a permanent salary increase. As far as I am aware, 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 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 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, consistent with 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 aligns closely with the empirical literature estimating discount factors (Best et al. 2018; Ericson and Laibson 2018).[17] This helps support the claim that teachers respond realistically to the experiment.

---

[16] Specifically, each coefficient in Deck 2 is multiplied by $\beta_{salary}^{Deck1}/\beta_{salary}^{Deck2}$, relating estimates across decks to be in the same units. Each coefficient in Deck 3 is multiplied by $\beta_{salary}^{Deck1}/\beta_{salary}^{Deck3}$.

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

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. Teachers in these schools already know their VA and principal evaluations from internal assessments, so the preference for performance pay in excess of its expected value may reflect its perceived nonpecuniary effects. 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 less costlessly (Phipps 2018). While a teacher can prepare for a small number of scheduled observations, success in value-added models (VA) requires continual effort. Alternatively, teachers may prefer an objective measure to an observation score that could be permeated by bias of evaluators. 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 their 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 reviewed 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, suggesting 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 equivalent of 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. 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, suggesting weaker preferences. 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 cost effective at improving 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’’ of 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 regarding disruptive students, I contacted a random sample of respondents, who indicated that a supportive principal would meet with disruptive students, support the teacher in enforcing discipline, and side with the teacher in disputes over discipline with parents.

An important question is whether supportive principals reduce teacher aversion to working in low-income or low-achieving 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 in mitigating those costs.

IV. Using Compensation to Affect Selection

In this second step of the paper, I examine the scope for compensation and working conditions to affect selection. If excellent teachers have distinctive preferences, a structure that differentially appeals to them can improve the distribution of teacher quality by altering selection.

Whether compensation and working conditions can affect selection depends on whether excellent teachers have distinctive preferences. For instance, 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; they may place a greater value on high starting salaries if they have more competitive outside options; or excellent instructors may have a distinctive appreciation for generous pensions if they are more committed to a long career in teaching. It’s also important to know whether highly rated teachers have different preferences for working conditions that are not affected by policy (like student demographics) to understand the payment levels needed to draw high-performing teachers into low-income schools.

To measure teacher performance, I estimate value-added models (VA) from student data and incorporate Danielson observation scores. The student 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 running from 2011 through 2016. I estimate VAs using a flexible polynomial of past scores, student fixed effects, school-year fixed effects, and indicators for whether last year's test score is missing in each subject (see online Appendix A). The VA model implies that a 1 standard deviation (SD) improvement in teacher VA raises normalized test scores by 0.16 SD in math and 0.07 SD in reading, similar to Chetty et al. (2014a). The VA used in the primary analysis is the average of a teacher's math and reading value-added. The resulting VA measure is 0 on average with a standard deviation of 1. I sort teachers into ten deciles based on their VA and generate a performance index from those deciles from 0 to 1. Since students are not tested in all grades, there are records to estimate value-added for half of teachers. To provide a measure of quality that covers a broader array of teachers, I supplement with Danielson observation scores which reflect yearly 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) and assign deciles based on the total score to generate a quality index from 0 to 1. The VA index and the Danielson index are significantly correlated for teachers with both measures ($p < 0.001$). For those teachers who do not have a VA index, I input the Danielson index as their quality measure. Together, the VA index and the Danielson index provide a quality measure for about 80 percent of respondents. I find the same results when using either measure in isolation.[18]

To test whether preferences vary by teacher rating, I interact each of the attributes from table 2 with the quality index in a model of teacher choice. To show visually how preferences vary throughout the teacher-quality distribution, I interact decile dummies with each attribute and plot the resulting coefficients. In both the statistical test and the nonparametric figures, 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 distinct preferences related to experience and not necessarily their ability to teach. The results are also robust to controlling for experience bins interacted with each attribute (table 4).

[18] See online Appendix table 5.

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 $160 salary increase. Teachers in the top decile value the same merit program as equivalent to a $610 salary increase (the interaction $p < 0.001$).[19] If teachers entertained two comparable offers, a high-performing teacher (top decile) is 22 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.

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 9 and 10, however, have significantly stronger preferences for performance pay than low-decile teachers. The top decile is 4.1 percent ($p = 0.010$) more likely to select an offer providing $1,000 in merit pay and teachers in the next top decile are 3.7 percent ($p = 0.004$) more likely. Corollary plots for each of the other attributes lack a systematic pattern, findings that are consistent with the results in table 4 and in online Appendix tables 3 and 4 in which higher quality teachers do not differ significantly in their preferences for school attributes. In future work, it may be fruitful to examine whether there are differential preferences for other attributes including dismissal rules and measures of colleague quality.

The Preferences of Marginal Teachers

[19] 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, 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.

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, and they may also have preferences similar to marginal prospective teachers who, also being near indifference, choose not to become teachers. 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 largely identical preferences for compensation structure and student characteristics. Of the 18 attributes in the study, teachers who leave the profession have different preferences in two attributes, 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 preferences (online Appendix tables 6–8). What this suggests is that attrition is largely unrelated to preference heterogeneity for working conditions in the district. What differs must be tastes for teaching.

To examine whether the preferences of marginal teachers differ on entry, I survey 1,193 college students in a large public university near the district. Students are asked to report how likely they are to enter teaching (on a 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. What is of interest here is whether those planning on teaching have similar preferences to marginal teachers.

Comparing the preferences of those set on teachings 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 differ little along the teacher-propensity index. The joint significance of attributes interacted with the teacher-propensity index is *insignificant* in the compensation deck. 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 deeper aversion to larger classes and a stronger preference for supportive principals than those who do not intend to teach.

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 likely 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 (e.g., new teachers won't have distinct preferences that would imply a different compensation structure). The uniformity of compensation preferences along the interest-in-teaching dimension also implies that it is tastes for teaching that drives entry and exit, not different preferences for compensation structure among teachers.

The data potentially lend themselves to other heterogeneity analyses. Though it is not directly useful to the aim of this paper, I present preference heterogeneity by sex, race, experience, and grade range in online Appendix C.

V. Optimizing the Compensation Structure

In this final step of the paper, I take the preference estimates at face value and calculate policies that would advance various school objectives, principally maximizing student achievement.

Compensation Structure

What do preferences suggest about how schools "should" structure compensation? Subject to the current budget constraint, I calculate the structure of teacher compensation that maximizes three related objective functions schools might pursue: First, I consider an objective that allocates resources to maximize the utility of teachers. Second, I calculate the compensation structure to maximize teacher retention. Third, I calibrate a model of the schools' achievement production function from the literature including the influence of teacher experience (Papay and Kraft 2015), class size (Krueger 1999; Hoxby 2000; Cho Glewwe, and Whitler 2012), and performance pay (Imberman and Lovenheim 2015). Retention—which gives rise to experience—is influenced by the teacher utility from compensation and working conditions. Performance pay affects achievement by eliciting effort from teachers and by retaining high performers (Lavy 2002, 2009; Imberman and Lovenheim 2015; Biasi 2021; Lazear 2000, 2003). I use the utility estimates from my experiment to simulate quality-specific attrition patterns as performance pay increases, allowing me to calculate the resulting distribution of teacher VA from introducing various levels of performance pay.

All simulations are based on the same model of teacher utility derived from my field experiment. By using the estimated utility function for current teachers, I assume that incoming teachers have similar preferences and the same quality distribution. The first assumption is supported by our examination of preferences along the propensity-to-enter-teaching index. The second assumption produces a conservative estimate if performance pay also encourages positive selection on entry. The optima in some exercises fall outside of the experimental range. Since preferences are primitives (and not treatment effects) 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 schools might maximize. The first is simply an objective to maximize teacher utility. This may be the goal of a district with a strong union presence that faithfully 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. I include starting pay as a linear numeraire (online Appendix tables 9 and 10).

$$U_a = (\widehat{\beta_1^1}\, S_a + \widehat{\beta_2^1} G_a + \widehat{\beta_3^1} G_a^2 + \widehat{\beta_4^1} P_a + \widehat{\beta_5^1} P_a^2 + \widehat{\beta_6^1} M_a + \widehat{\beta_7^1} R + \widehat{\beta_8^1} R^2 + \widehat{\beta_9^1} D + \widehat{\beta_{10}^1} H)/\widehat{\beta_1^1} + (\widehat{\beta_2^2} C + \widehat{\beta_3^2} C^2)/\widehat{\beta_1^2} \qquad (2)$$

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*). The equation blends utility estimates on compensation from the compensation-structure deck and utility estimates on class size from the working-conditions deck. Without allowing for nonlinearity, the results would degenerate to a corner solution in which all compensation would load 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 by dividing each of the coefficients by the beta on starting salary (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 of utility in this objective will refer to the average utility of faculty (Hoxby 1996; Figlio 2002).

Maximizing retention (the average experience of teachers) is a related objective, and teacher experience is one of few reliable predictors of teacher performance (Wiswall 2013). Hendricks calculates base retention probabilities over the life cycle of a teacher as well as how those probabilities change in response to an exogenous increase in salary (Hendricks 2014). These estimates neatly cohere with my utility calculations since they are also in units of salary. Because the estimates in Hendricks (2014) come from Texas, they likely generalize to teachers in 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:

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

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. Suggesting the reliability of the model, the average tenure predicted using status quo compensation structure matches the district's actual average experience (9.0 years). To determine the average tenure, I calculate the share of teachers remaining in each experience cell to simulate the *equilibrium* experience distribution: the stock of teachers in experience cell *e*, is simply the number remaining from the 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 $\mathbf{D}_{\mathrm{e}} = [D_1, D_2, \ldots, D_{35}]$ where D_e states the share of teachers employed with experience level *e* in equilibrium. The object I maximizes is the *average* level of teaching experience for allocation *a*:

$$E_a = \sum_{e=1}^{35} D_e \times e \tag{4}$$

The primary objective I consider is the maximization of student achievement. The model incorporates several well-studied features of the production function. Students learn more in smaller classes (Krueger 1999; Hoxby 2000; Cho, Glewwe, and Whitler 2012). Improving teacher welfare affects the retention probabilities in each experience cell. Retention increases achievement since more experienced teachers have increasing, concave impacts on student learning (Papay and Kraft 2015). To calculate the influence of experience on achievement, I calculate retention probabilities, as above, simulate the equilibrium experience profile, and take the dot product of the experience distribution with VA-gains-by-experience vector from Papay and Kraft. Performance pay improves achievement by eliciting additional effort (Lavy 2009; Imberman and Lovenheim 2015), and produces positive selection in retention based on preferences.

To calculate the influence of performance pay on selection, I take a large cross section of simulated new teachers, uniformly distributed in performance. I calculate individual utility based on the attribute bundle with heterogeneity in preferences along the performance dimension. I add to their calculated utility a random component from the empirical distribution of the error terms in preference model. This accounts for the fact that many of the factors affecting teacher choice are outside of the empirical model. Without incorporating the random influence of unobserved factors, only the highest performing teachers would remain. After calculating the quantity who exit each year from the retention probabilities r_{ea}, I remove teachers with the lowest utility up to that cutoff and repeat the process for every experience cell over the life cycle. The result produces the equilibrium "quality" distribution, which I denote $\mathbf{Q}_{\mathrm{d}}$, where each Q_d describes the share of teachers in equilibrium who are in the d^{th} decile of the initial performance distribution. In practice, re-solving for this distribution each time the search iterates is computationally burdensome. I linearize the problem by calculating the quality distribution with no performance pay, and also the quality distribution with $4,000 in performance pay and calculate the average change in value-added for a $1,000 increase in performance compensation. I allow the gradient to differ when performance pay is based on value-added models alone or in conjunction with observation scores. Teachers prefer to be evaluated on both, but pay based on value-added models alone more closely targets higher payments (and retention) to achievement-enhancing teachers.

The resulting achievement production function averages the per-student impact of class size changes in domestic studies across grades (Krueger 1999; Hoxby 2000; Cho, Glewwe, and Whitler 2012). A thousand dollars in performance pay increases achievement via effort by 0.014 standard deviations from Imberman and Lovenheim (2015), whose study has the virtue of being from the same setting (Houston, Texas) and thus likely generalizes to the district. Performance pay increases teacher quality through selection in equilibrium by 0.017 standard deviations if it is based on value-add and observation measures; if it is based on value-added models alone, teacher utility is somewhat lower, but teacher quality rises by 0.023 standard deviation from selection for an additional $1,000 in performance pay. The effect of teacher retention is the dot product of the experience distribution and experience-specific value-added measures times the average achievement gains for a 1 SD increase in teacher quality (0.115, the average student improvement (averaging math and reading) from a 1 SD improvement in teacher VA).

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

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

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 school has to pay a fraction of salary ϕR to retirement accounts. Therefore the cost parameter ϕ reflects the needed contribution for a one-percent replacement in retirement. There is a budget cost to turnover t (Barnes, Crowe, and Shaefer 2007; Watlington et al. 2010). Retention is therefore budget 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 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 the 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. Class-size reductions become more costly as total compensation rises since hiring additional teachers become more expensive.

I constrain optimization to conform to some practical requirements. No unit of compensation can be negative. I've included class size as a way of seeing whether smaller classes function as a cost-effective compensation provision to teachers or a cost-effective means of achievement promotion, and so I constrain class size so that it cannot rise greater than 30 students per class (the status quo is 29). Without this constraint, the model pushes towards large classes with very well-paid teachers. Performance pay is also constrained so that it can be no larger than \$5,000. Without this constraint, some models recommend substantial allocations of performance pay. 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 go into further detail about the maximization exercise and cost calibration in online Appendix D.[20]

[20] Inattention in the survey will suggest a larger random component than exists in nature. If inattention played a role, the achievement effects discovered in the simulation will tend to be conservative.

I solve the optimization problem using a nonlinear programming solver. For inference, I bootstrap 1,000 estimates of teacher utility and solve the maximization problem separately with each estimate to produce an empirical distribution of optima consistent with the data. Results are presented in Table 5. Stars indicate an attribute is statistically distinct from the status quo allocation.

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 promised to replace 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 in retirement (20 percent). Schools would also shift to defined-contributions retirement plans that are both less costly to districts and more attractive to teachers. 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 using the other objective functions specified in the last section. Maximizing teacher utility would increase teacher retention and thereby raise average teacher experience by 21 percent in equilibrium. This reform also increases student achievement by 0.044σ each year, which comes from increased teacher experience (47%), induced effort from merit pay (46%), and improved retention of high-caliber teachers (7%).

Moving to a defined-contributions plan may not be politically feasible. 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 new evidence suggests that teacher output improves throughout a career (Wiswall 2013; Papay and Kraft 2015). Whenever any teacher departs, it opens a vacancy chain that causes an additional novice to be hired somewhere, which reduces achievement.

The compensation structure that maximizes retention implies starting salary above the status quo ($66,688). The optima targets higher compensation to teachers that already have experience with a positive salary growth rate of 1.4 percent. Like the teacher-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 (3.5 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.[21]

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%).

Compensation Structure to Maximize Student Achievement

Improving teacher welfare may not directly increase achievement (for example, de Ree et al. 2018). The reform that maximizes achievement would include higher base pay than the status quo ($66,774), a modest rate of salary growth (1.3 percent growth rate), $5,000 payments in performance pay, and a slightly larger class size (3.5 percent). Whereas the other optimizations recommended using VA in combination with observation scores to distribute performance pay, this model recommends using VA-only to evaluate performance. This practice is not preferred by teachers, but it improves targeting payments to high-VA teachers to reduce their attrition. The

[21] 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, but they do little to retain younger teachers. These simulations suggest that defined contributions plans increase teacher tenure, 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 most teachers in DC plans than in pensions.

resulting achievement-optimal bundle reduces the replacement rate by 17 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 take form over decades. 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 model predicts partial-equilibrium effects where a single district reforms its compensation structure. How would the effect scale in general equilibrium, where all districts adopt the reform? The key is to understand the degree to which local compensation-induced retentions prevent (i) transfers to other districts or (ii) departures from the profession. The former retentions come at the expense of other districts and disappear in general equilibrium. The latter do not.[22] I estimate the effect of compensation on district transfers and professional departure using administrative staffing records from Texas and a triple-difference design following Hendricks (2014). The triple-difference leverages variation within the district across experience cells when districts change their salary schedules. One-third of those retained by local-salary increases were dissuaded from district transfers while two-thirds were dissuaded from leaving the profession.[23] This suggests districts have significant market power as employers. I assume that the direct effects of class size and effort scale in general equilibrium. The results suggest that the general equilibrium effect on achievement would be 10 percent smaller than the partial equilibrium effect. The approach and results are described in more detail in online Appendix E.

It is not clear from this examination whether performance pay generates positive selection into teaching. Though the question is beyond the scope of this study, two testable conditions are necessary. First, prospective teachers would need to have private information regarding their ability to teach before they enter the profession. If the beliefs of prospective teachers about their effectiveness is uncorrelated with their eventual quality, performance pay programs will fail to drive positive selection on the entry margin. Second, marginal teachers must have similar

[22] I assume the direct effect of class size and the effect of incentives on effort remain in general equilibrium.
[23] 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.

(affirmative) preferences for performance pay as other teachers. Both in the district and among prospective teachers, marginal teachers have similar preferences for performance pay.

Across objectives, the maximization exercises suggest 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 provide an indication for whether the district compensation structures are distorted from its own optimal based on those already there. It is striking that—even among a selected group of teachers choosing the district—the status quo compensation structure does not appear to reflect either teacher preferences or a structure that maximizes tenure 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. That the 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 teacher welfare, teacher retention, or achievement maximization. It may be that political constraints or bargaining affect compensation structure. Since unions are typically led by older, veteran teachers, they might bargain for compensation structures that reflect their private preferences, especially if the costs of pensions are shrouded to voters (Glaeser and Ponzetto 2014).[24] If true, we would expect places with stronger union presence to pay a larger share of compensation in benefits.[25] I gather a measure of state-level union strength provided by the Fordham Institute, which identifies the strength of unions based on five measures: resources and membership, involvement in politics, scope of bargaining, state policies, and perceived influence. These factors are combined to form five quintiles, with the top quintile representing states with the strongest union presence. A one quintile increase in union strength is associated with a benefit-share increase of 2.6–2.8 percentage points

[24] 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.

[25] There is a strong negative relationship between total compensation and salary share. There is also a strong relationship between total compensation and union strength. I control for total compensation to avoid confounding benefit-share increases with increased total compensation.

($p < 0.001$), explaining a 10-point difference between states with the weakest unions (where compensation is 29.8 percent benefits) and where unions are strongest (where compensation is 39.8 percent benefits), conditional on total compensation (online Appendix table 12).

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.[26] 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. 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 school 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 preferences for teaching among college students using a similar hypothetical choice design. This would further illuminate how to efficiently draw able educators into the 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 recruit and retain excellent instructors. And, considering the apparent importance of principals, 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 attributes that don't vary in the real world. In practice, competing schools offer the same compensation and contracts because of market concentration, pattern bargaining, and public regulations. Third, in addition to

[26] 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.

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 realistic 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

Alpizar, Francisco, Fredrik Carlson, and Olof Johansson-Stenman. “Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica.” *Journal of Public Economics* 92, no. 5-6 (2008): 1047-1060.

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.

Atkin, Charles K., and Steven H. Chaffee. “Instrumental Response Strategies in Opinion Interviews.” *Public Opinion Quarterly* 36, no. 1 (1972): 69-79.

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.

Beggs, Steven, Scott Cardell, and Jerry Hausman. “Assessing the Potential Demand for Electric Cars.” *Journal of Econometrics* 17, no. 1 (1981): 1-19.

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. “Unions, Salaries, and the Market for Teachers: Evidence from Wisconsin.” *Working paper* (2019).

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).

Bulte, Erwin, Shelby Gerking, John A. List, and Aart De Zeeuw. “The Effect of Varying the Causes of Environmental Problems on Stated WTP Values: Evidence from a Field Study.” *Journal of Environmental Economics and Management* 49, no. 2 (2005): 330-342.

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.

Camerer, Colin F., and Robin M. Hogarth. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19, no. 1-3 (1999): 7-42.

Campbell, Bruce A. "Race-of-Interviewer Effects among Southern Adolescents." *Public Opinion Quarterly* 45, no. 2 (1981): 231–244.

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.

Cotter, Patrick R., Jeffrey Cohen, and Philip B. Coulter. "Race-of-Interviewer Effects in Telephone Interviews." *Public Opinion Quarterly* 46, no. 2 (1982): 278–284.

Cummings, Ronald G., and Laura O. Taylor. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89, no. 3 (1999): 649-665.

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.
Finkel, Steven E., Thomas M. Guterbock, and Marian J. Borg. “Race-of-Interviewer Effects in a Pre-Election Poll Virginia 1989.” *Public Opinion Quarterly* 55, no. 3 (1991): 313-330.
Fisher, Robert J. “Social Desirability Bias and the Validity of Indirect Questioning.” *Journal of Consumer Research* 20, no. 2 (1993): 303–315.
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.
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.
Green, Paul E., and Venkat Srinivasan. “Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice.” *The Journal of Marketing* (1990): 3-19.
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.

Herriges, Joseph, Catherine Kling, Chih-Chen Liu, and Justin Tobias. "What Are the Consequences of Consequentiality?" *Journal of Environmental Economics and Management* 59, no. 1 (2010): 67-81.

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.

Karlan, Dean S., and Jonathan Zinman. "List Randomization for Sensitive Behavior: An Application for Measuring Use of Loan Proceeds." *Journal of Development Economics,* 98, no. 1 (2012): 71-75.

Kofoed, Michael S. and Elizabeth McGovney. "The Effect of Same-Gender and Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point." *Journal of Human Resources* (2017): 0416-7838r1.

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).

Krosnick, Jon A. "Maximizing Questionnaire Quality." *Measures of Political Attitudes* 2 (1999): 37-58.

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.

Leggett, Christopher G., Naomi S. Kleckner, Kevin J. Boyle, John W. Dufield, and Robert Cameron Mitchell. "Social Desirability Bias in Contingent Valuation Surveys Administered Through In-Person Interviews." *Land Economics* 79, no. 4 (2003): 561-575.

List, John A. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review* 91, no. 5 (2001): 1498-1507.

List, John A. "Does Market Experience Eliminate Market Anomalies?" *The Quarterly Journal of Economics* 118, no. 1 (2003): 41-71.

List, John A. "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace." *Econometrica* 72, no. 2 (2004): 615-625.

List, John A., Robert P. Berrens, Alok K. Bohara, and Joe Kerkvliet. "Examining the Role of Social Isolation on Stated Preferences." *American Economic Review* 94, no. 3 (2004): 741-752.

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 Willén. "The Long-run Effects of Teacher Collective Bargaining." *American Economic Journal: Economic Policy* 11, no. 3 (2019): 292-324.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan L. Weigel. "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom." *Econometrica* 85, no. 4 (2017): 1065-1091.

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.
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., 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).
Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. "Truth in consequentiality: theory and field evidence on discrete choice experiments." *American Economic Journal: Microeconomics* 4, no. 4 (2012): 145-71.
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.
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.

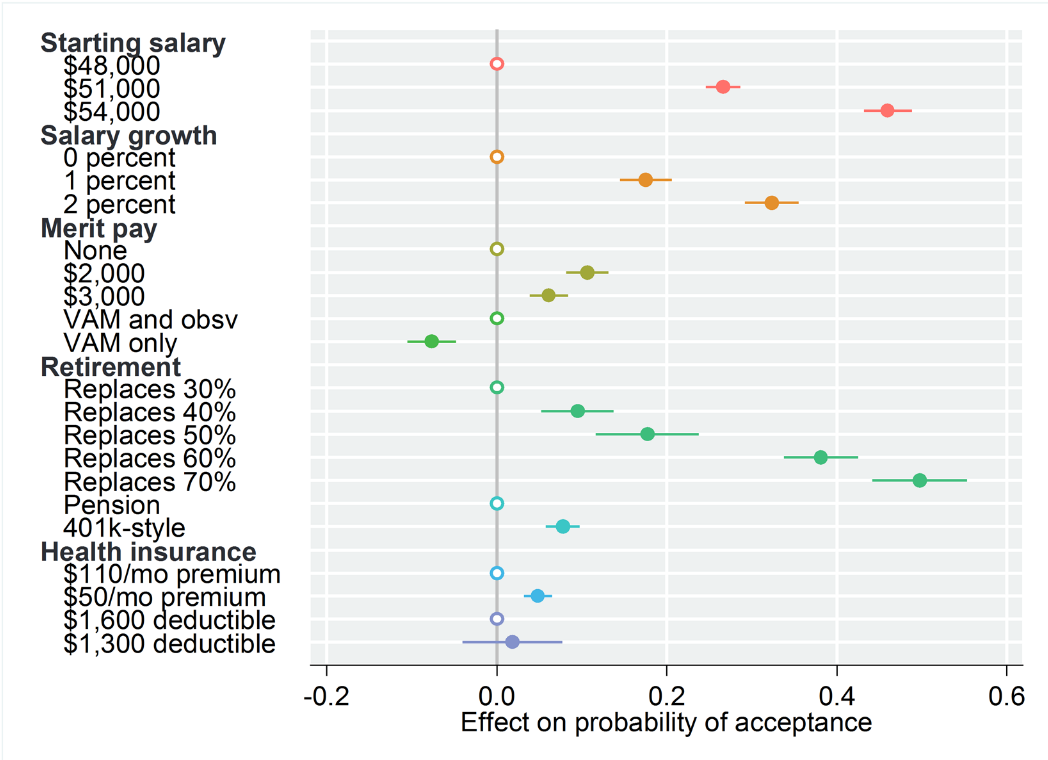
FIGURE 1—SAMPLE QUESTION

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

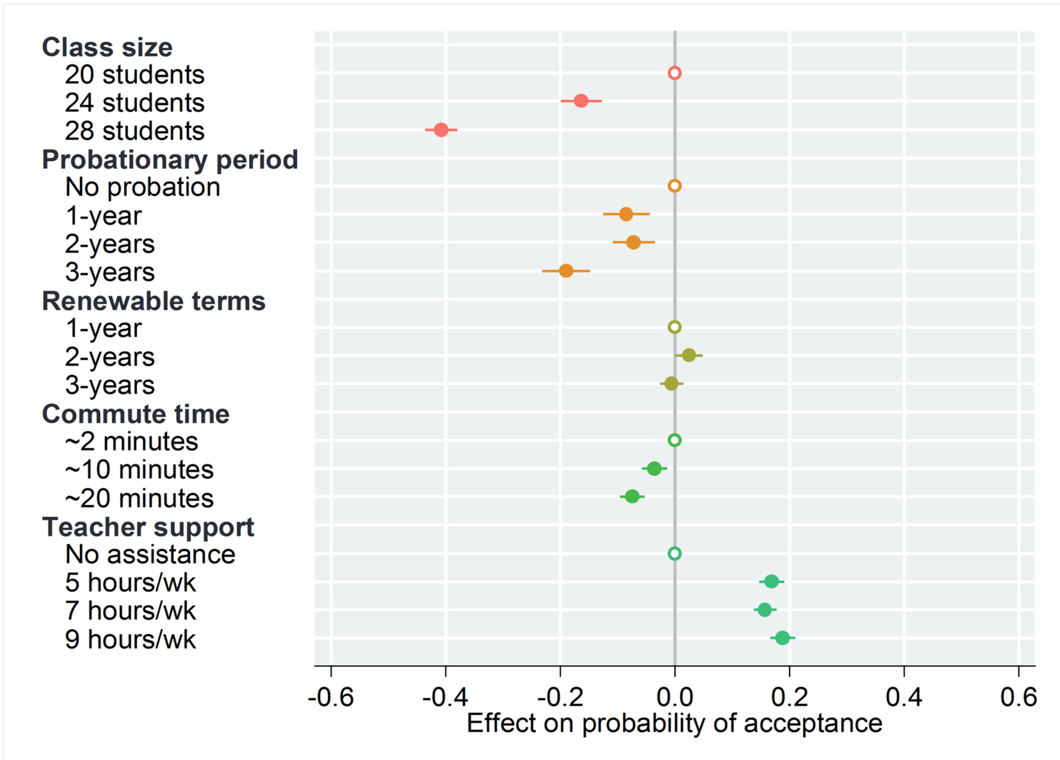
Note: This figure presents an illustration of the questions answered by teacher respondents about compensation structure.

FIGURE 2—EFFECTS OF COMPENSATION ATTRIBUTES
ON THE PROBABILITY THAT TEACHERS ACCEPT A JOB OFFER



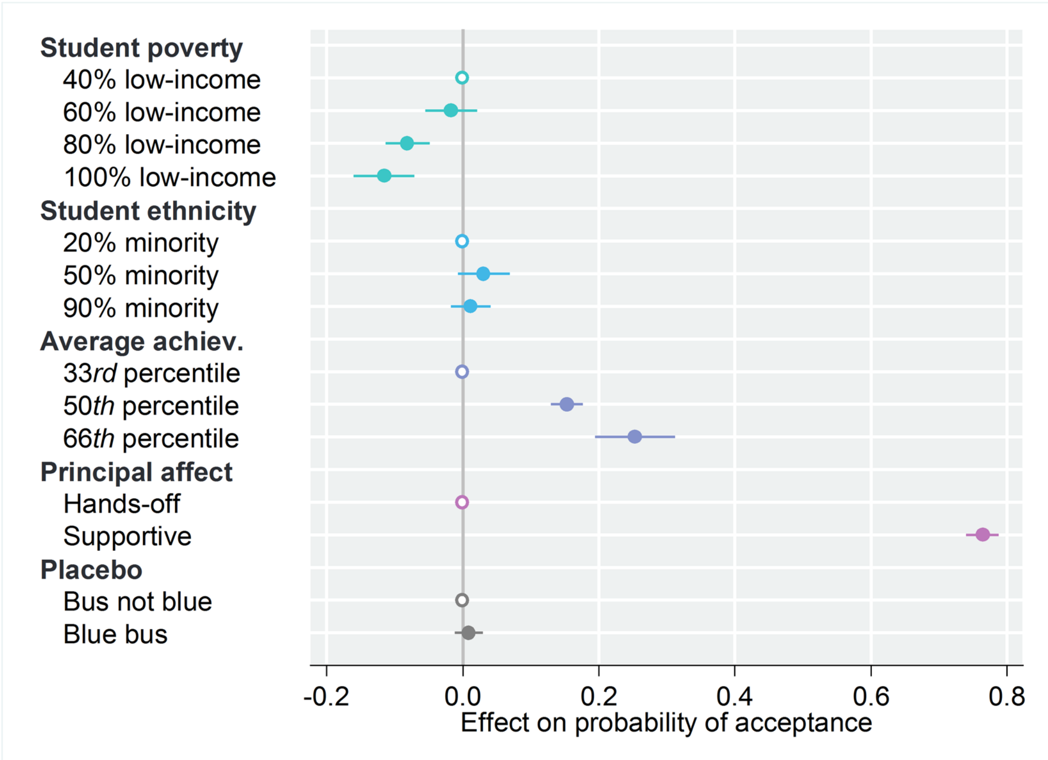
Note: Dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from least-squares regression. The unfilled dots on the zero line denote the reference category for each job-offer attribute.

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



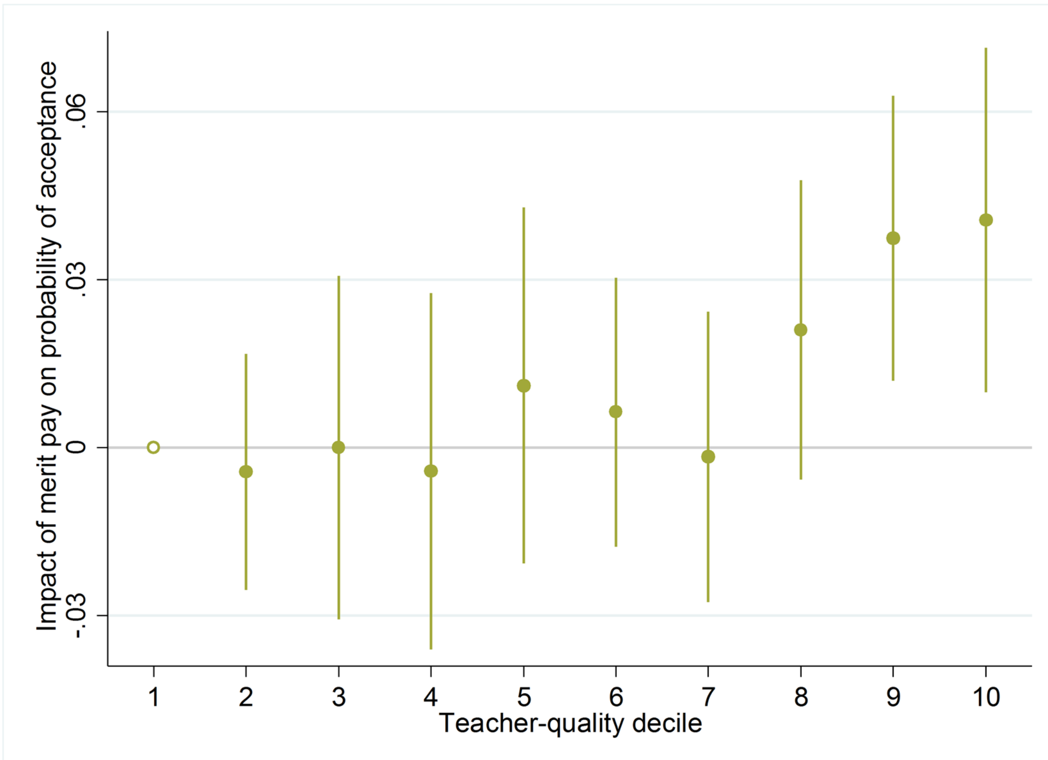
Note: Dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from least-squares regression. The unfilled dots on the zero line denote the reference category for each job-offer attribute.

FIGURE 4—EFFECTS OF STUDENT AND PRINCIPAL ATTRIBUTES ON THE PROBABILITY THAT TEACHERS ACCEPT A JOB OFFER



Note: Dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from least-squares regression. The unfilled dots on the zero line denote the reference category for each job-offer attribute.

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



Note: In this figure, I identify the teacher-quality decile of each teacher using VA and, for those teachers who lack a VA score, the decile of their Danielson observation score. The coefficients above represent the differential effect of merit pay (in $1,000s) on the probability a teacher will accept a job offer.

TABLE 1—SUMMARY STATISTICS ON OFFER ATTRIBUTES FOR CONJOINT EXPERIMENT

	Average	Std. Dev.	Units
Choice	0.50	(0.50)	Indicator
Starting Salary	49.51	(2.38)	$1,000s
Salary Growth	1.44	(0.71)	% growth
Bonus amount	1.25	(1.29)	$1,000s
VA only	0.20	(0.40)	Indicator
Replacement	48.09	(9.31)	% of salary
401k-style	0.37	(0.48)	Indicator
Premium (yearly)	0.78	(0.30)	$1,000s
Deductible	1.48	(0.18)	$1,000s
Probationary period	1.72	(0.93)	Years
Term length	2.26	(0.96)	Years
Commute time	0.187	(0.096)	Hours
Class size	24.55	(3.39)	Students
Assistance	3.26	(3.66)	Hours/week
Percent low income	6.79	(1.86)	10%s
Percent minority	5.62	(2.97)	10%s
Ave. achievement	4.99	(1.65)	10%tiles
Supportive	0.42	(0.49)	Indicator
Blue bus	0.50	(0.50)	Indicator

Note: This table presents the mean and standard deviation of the experimental data. The units column describes the units of each variable to aid interpretation of regression results.

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	0.395**	(0.008)	$1,000
Salary growth	0.192**	(0.009)	$2,270	0.948**	(0.034)	$2,400
Merit reward						
Bonus amount	0.029**	(0.003)	$346	0.192**	(0.012)	$486
VA only	-0.077**	(0.015)	-$907	-0.209**	(0.055)	-$529
Retirement						
Replacement	0.015**	(0.001)	$173	0.071**	(0.002)	$181
401k-style	0.077**	(0.010)	$907	0.413**	(0.035)	$1,046
Health insurance						
Premium (yearly)	-0.082**	(0.014)	-$970	-0.438**	(0.048)	-$1,109
Deductible	-0.312	(0.212)	-$3,688	-1.009	(0.760)	-$2,554
Panel 2: Working-Conditions Deck						
Contract						
Probationary period	-0.058**	(0.005)	-$502	-0.320**	(0.022)	-$467
Term length	-0.004	(0.005)	-$33	0.014	(0.021)	$21
Working conditions						
Commute time	-0.365**	(0.043)	-$3,177	-2.880**	(0.200)	-$4,204
Class size	-0.068**	(0.001)	-$595	-0.399**	(0.007)	-$582
Assistance	0.030**	(0.001)	$257	0.175**	(0.005)	$255
Panel 3: Students-&-Leaders Deck						
Students						
Percent low income	-0.022**	(0.002)	-$324	-0.117**	(0.010)	-$285
Percent minority	0.0027	(0.0014)	$40	0.007	(0.006)	$18
Ave. achievement	0.036**	(0.003)	$546	0.237**	(0.011)	$577
Principal affect						
Supportive	0.575**	(0.009)	$8,673	3.04**	(0.042)	$7,392
Placebo						
Blue bus	0.007	(0.008)	$101	0.019	(0.038)	$47

Notes: * $p < 0.05$, ** $p < 0.001$. Each coefficient represents the parts worth impact of an attribute on the odds of accepting a presented job offer. These estimates are translated into willingness-to-pay values by scaling the impact of an attribute by the impact of $1,000 starting salary. 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)
Principal supportive (PS)	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 × PS	.	-0.045**	-0.061**
	.	(0.011)	0.0115
Poverty rate	-0.022**	-0.020**	-0.033**
	(0.002)	(0.003)	(0.005)
Poverty × PS	.	.	0.028*
	.	.	(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 presents the results of linear probability models in which I test whether having a principal "supportive with disruptive students" attenuates a teachers' aversion to poorer or lower-achieving school settings.

TABLE 4—TEACHER PREFERENCES BY QUALITY

	Choice		Choice	
	Reference Group (1)	Quality-index interaction (2)	Reference Group (3)	Quality-index interaction (4)
Salary				
Starting salary	0.090**	-0.002	0.091**	-0.001
	(0.004)	(0.006)	(0.004)	(0.006)
Salary growth	0.178**	0.004	0.183**	0.008
	(0.014)	(0.017)	(0.014)	(0.017)
Merit reward				
Bonus amount	0.014*	0.041**	0.018*	0.041**
	(0.007)	(0.011)	(0.007)	(0.011)
VA only	-0.064*	-0.025	-0.075*	-0.022
	(0.022)	(0.027)	(0.025)	(0.028)
Retirement				
Replacement	0.013**	0.002	0.013**	0.002
	(0.001)	(0.0014)	(0.001)	(0.0014)
401k-style	0.062*	0.034	0.079**	0.042
	(0.019)	(0.030)	(0.022)	(0.030)
Health insurance				
Premium (yearly)	-0.112**	0.071	-0.106**	0.071
	(0.031)	(0.054)	(0.031)	(0.054)
Deductible	-0.453	-0.130	-0.270	-0.163
	(0.284)	(0.226)	(0.287)	(0.225)
Experience bins	X		X	
Exp. interactions	.		X	
R-squared	0.201		0.203	
Observations	21,358		21,358	

Note: * p < 0.05, ** p < 0.001. Columns (1) and (2) represent one regression in which the main effects are displayed in column (1) and the interactions with the quality index are represented in column (2). The regression displayed in columns (3) and (4) follows a similar form but adds controls for experience bins interacted with each attribute.

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**	\$66,774**
Salary growth	1.8%	0.0%**	1.4%	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	79.2	**96.3**	90.8	85.0
Teacher experience	9.03 years	10.9 years	**11.0 years**	10.0 years
Student achievement	0.092σ	0.135σ	0.136σ	**0.183σ**

Note: * p < 0.05, ** p < 0.001. This table presents the results of maximizing teacher utility, teacher experience, and student achievement subject to the current budget constraint. Statistical significance is calculated by bootstrapping 1,000 estimates of the utility function and re-solving the maximization problem for each one.