Preferences, Selection, and the Structure of Teacher Compensation

Andrew C. Johnston[1]

February 18, 2021

Abstract

Human capital shapes income, inequality, and growth. In the public sphere, human-capital formation depends largely on the selection and retention of teachers. To understand how to improve selection and retention, I use a discrete-choice experiment to estimate teacher preferences for compensation structure, working conditions, and contracts. High-performing teachers have stronger preferences for schools offering performance pay, which implies it promotes positive selection. Under a variety of school objectives, schools appear to underpay in salary and performance pay while overpaying in retirement. The results suggest significant efficiency gains from restructuring compensation: teacher welfare and student achievement can be simultaneously much improved.

JEL Codes**:** I20, J32, J45, M50

[1] Johnston: Department of Economics, University of California, Merced; email: acjohnston@ucmerced.edu. For clear insight, I am grateful to David Card, Damon Clark, Mark Duggan, Laura Giuliano, Alex Mas, Jesse Rothstein, Kenneth Shores, 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 major factor shaping income, inequality, and growth (Neal and Johnson 1996; Barro 2001; Chetty, Friedman, and Rockoff 2014). In the formation of human capital, teachers are quite possibly the most influential 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 in adulthood (Chetty et al. 2011; Petek and Pope 2019).[2] Merely replacing a poor teacher with a median one for a single year is estimated to be worth over half a million dollars in students' future earnings in net present value (Chetty, Friedman, and Rockoff 2014; Opper 2019; Gilraine and Pope 2020).[3] The effect teachers have on human capital varies widely and improves systematically with experience. What this implies is the focal importance of teacher *selection* and *retention* in building the stock of human capital.

Governments typically play centrally in the provision and financing of education. In the United States, governments spend almost $1 trillion per year on K–12 education, the principal cost of which is *personnel*. Teachers are subject to a distinctive compensation structure: low salary, generous retirements, and no performance incentives. Because public schools operate as local monopolists with significant market power, it is not obvious that schools will structure compensation to maximize achievement (Hoxby 2000; Rothstein 2007; de Ree et al. 2018). Political incentives, in particular, may distort compensation structures 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 massive investments in teacher compensation are structured *well.*

Teacher compensation can affect human capital formation through a few channels. If compensation is allocated to deliver the most utility for teachers (subject to a budget constraint), the value of teaching rises and the potential for retention improves. By delivering teachers a preferred package, schools can deliver students more experienced, effective faculty. Likewise, if high-performing teachers have distinctive preferences, simply structuring compensation to appeal

[2] For example, Chetty et al. (2014) find that being exposed to a teacher with 1σ higher VAM for a single year increases a student's future earnings by about 1 percent each year; these students are also more likely to attend college, less likely to have children while in high school, and they save a greater share of their income. What's more, the impact of teaching extends beyond the students in their classroom. Peer effects from teaching quality suggest the ultimate effect of a quality teacher is about a third larger than typical VA measures (Opper 2019). Gilraine and Pope (2020) find that traditional VA measures contain substantial measurement error driven by short-term improvements. VA measures that avoid mismeasurement imply that teacher quality is *twice* as influential as previously measured.

[3] Providing talented teachers is a rare intervention that produces long-term benefits, especially for low-income children. See, for instance, Altonji and Mansfield (2011); Dahl, Kostol, and Mogstad (2014); Chetty, Friedman, and Rockoff (2014); Heckman, Humphries, and Veramendi (2018).

to high-performers can induce positive selection. In these channels, compensation affects achievement by guiding teachers' labor supply on the extensive margin. In some cases, compensation directly affects the *intensive* margin in a way that influences student achievement. Performance pay elicits greater effort by teachers, and reductions in class size focus a teacher's attention on fewer students.

I examine the ideal structure of teacher pay in three intuitive movements. First, I estimate teacher preferences for each component of compensation and working conditions in a large urban district. Schools can improve the appeal of teaching by *reallocating* resources into compensation vehicles where the ratio between willingness-to-pay and cost-of-provision is high on the margin. Second, I test whether high-performing teachers have distinctive preferences that would be useful to *differentially* attract and retain them. Third, I use these estimates to calculate counterfactual compensation policies that achieve policy objectives subject to the current budget constraint. I compare the "optimal" bundles from each of the maximization exercises to what the district actually does to examine what goals schools actually pursue. I use the model to locate opportunities for efficiency gains: *how can policymakers reallocate pay to improve student achievement, and can they do so while also increasing* teacher *welfare*?

On the empirical front, teacher preferences are difficult to study. Normally, economists disentangle preferences by collecting data on the menu of options available when agents make a selection (Train 2009; Wiswall and Zafar 2017). The records necessary for this course (concurrent job offers) do not appear to exist, and teachers almost never entertain simultaneous offers.[4] Second, naturally occurring records would likely produce biased results since observed characteristics are endogenous to important unobserved ones. Most critically, the natural variation in attributes in the real world is extremely limited since contracts are essentially uniform and many important features are absent or colinear.[5]

To address these challenges, I deploy a discrete-choice experiment that permits me to estimate teacher preferences for compensation structure, contract type, and working conditions. I present primary- and secondary-school teachers with a series of hypothetical job offers, among which they

[4] Contacted districts did not keep records of job offers made. Conversations with firms that provide HR software to school districts indicate that fewer than 1% of schools use the software to make offers. Teachers, moreover, rarely entertain simultaneous offers because offers explode on the same day they are extended.

[5] State policy and common union influence generate similar compensation structures across districts. Within district, compensation is totally uniform. Many states provide a broadly shared pension and health insurance programs, rendering teacher choice uninformative. Importantly, real-world data are particularly unhelpful in determining preferences for merit pay or alternative retirement vehicles which almost never vary. When studying choices across states, say in a city that spans two states like St. Louis, the transition cost associated with state licensing may be such that teachers are only able to choose across state lines at an additional cost, collinear with any state-level differences.

select their preferred option. In each question, teachers make tradeoffs between features including starting salary, retirement generosity, larger merit rewards, smaller class sizes, principal support, and expedited time-to-tenure. The response rate is high (>97 percent). Their choices in the experiment illuminate the structure of their preferences. By estimating preferences over several facets of the work setting, I can assess the *allocation* of payments and calculate the "preferred" compensation structure, something not possible by a piecemeal approach.

There is good reason to believe the experimental design reveals true preferences. When discrete-choice experiments are tested, analysts find that hypothetical choices produce the same preferences as choices in the real world (Camerer and Hogarth 1999; Mas and Pallais 2017; Wiswall and Zafar 2018; Maestas et al. 2018). As examples, preferences estimated from hypothetical and real choice are *indistinguishable* (Mas and Pallais 2017) and hypothetical career choices accurately predict the eventual careers students select (Wiswall and Zafar 2018). Several features are conducive to truth-telling: subjects face tradeoff, discrete choices with many factors circumvent social desirability bias, and teachers have experience in the labor market for teaching. In this setting, moreover, teachers have a consequential reason to reveal their preferences: the survey was delivered to inform the district's new personnel regime, so their responses affect the new policy (Carson, Groves, and Machina 2000).

The experimental design elicits responses that match each realism benchmark available to me. For a handful of attributes, I can compare the estimates from this study to theory or touchstone literatures; consistently, the estimates retrieved here closely match these benchmarks, lending support to the other, more novel, estimates. If teachers pay part of their health insurance premium, they should be indifferent between an additional dollar of salary or an additional dollar offsetting what they pay for insurance. Reassuringly, teachers value health-insurance subsidies identically to an equivalent increase in salary. Similarly, the discount rate that rationalizes teachers' salary-retirement tradeoff is *exactly* that estimated in the empirical 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). The success on these benchmarks instills confidence that teachers responded realistically.

Policymakers can improve the appeal of teaching by shifting compensation into vehicles that teachers prefer relative to their cost. To understand how teachers value different components of their workplace, I use the choice experiment to estimate willingness-to-pay (WTP) for each of several attributes. Teachers value a ten-student class-size reduction (in a class of 30) equal to a

$5,950 increase in salary (11.9 percent of base pay),[6] seven times less than the cost of such a reduction. Teachers consistently prefer riskier, though portable, defined-contribution retirement plans over a traditional pension. Teachers also value quicker tenuring: an additional year of probationary status is equivalent to a salary reduction of $500 (1 percent). Many of these results are novel, and I provide additional estimates on the WTP for a broad array of other school attributes including shorter commutes, administrative support, and different evaluation schemes. To my initial surprise, the attribute teachers most value (that is, that having the highest odds ratio in choice) is a principal who "supports them with disruptive students." Having such a principal is valued equal to a 17.3-percent increase in salary. A "supportive" principal also reduces teacher aversion to teaching in disadvantaged settings. 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.[7] The results imply that student misbehavior is taxing but that attentive principals greatly reduce those costs.

I explore whether high-performing teachers have distinctive preferences which policymakers could use to shape selection. Forecasting which (prospective) 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-type teachers have distinctive preferences for conditions controlled by policy, policy*makers* can construct a "separating equilibrium" by structuring compensation, contracts, and working conditions to conform to the preferences of high performers (Ballou 1996; Hanushek 2011).[8] Using value-added models and principal evaluations, I find that high-performing teachers have broadly similar preferences to their colleagues, except in one regard. Excellent teachers systematically prefer jobs that include the opportunity to earn performance pay. High-performing teachers (top decile) are 22 percent more likely than low-performing ones (bottom decile) to select an offer providing $3,000 in merit pay, which induces favorable selection *in retention*. It is less clear whether merit pay would affect sorting *into* the profession since prospective teachers may not know their teaching ability prior to entering.

I use the model of teacher preferences to examine how schools would structure pay to achieve various objectives. Those objectives include maximizing (i) teacher welfare, (ii) teacher tenure,

[6] Here, base starting pay is $50,000 for a new teacher without a master's degree.
[7] Said another way, student poverty and achievement matter much less to teachers in the presence of a supportive principle.
[8] Over time, the effect may be especially pronounced since the preferred compensation differentially retains high-performing teachers who also prefer work settings inhabited by other high-caliber colleagues (Feng and Sass 2016). 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).

and (iii) student achievement. To this end, I estimate teacher utility with diminishing marginal returns, map teacher utility to attrition decisions (Hendricks 2014), and calibrate an achievement production function (Krueger 1999; Papay and Kraft 2015; Imberman and Lovenheim 2015).

Whether maximizing teacher utility, teacher retention, or student achievement, I find that teachers are underpaid in salary as well as performance pay and are overpaid in retirement benefits. Restructuring what teachers are paid—*subject to the current budget constraint*—to maximize their utility generates a 21.6 percent increase in teacher welfare, the equivalent of a permanent $17,000 raise (without spending any additional money). This naturally extends the average tenure of teachers by 21 percent.[9]

Structuring pay to maximize teacher tenure increases starting pay and includes a modest growth rate to promote retention among already-experienced teachers. The resulting compensation structure raises the average teacher tenure by 22 percent and increases the odds of a student having a veteran teacher by 33 percent. When maximizing tenure, achievement increases by 0.07σ per year, generated by a more experienced workforce and the introduction of a modest performance-pay program which teachers value more than its cost.

Restructuring pay to maximize student achievement also increases salaries and performance pay. Over time, this structure differentially retains high performers such that students are 24 percent more likely to have a teacher from the top of the original performance distribution. The achievement-optimal structure improves achievement by 0.19σ per year, though the full effect takes form over time. The achievement gains are driven by better overall retention, fostering a more experienced faculty (5%), added effort by teachers (35%), and positively selected retention (60%). This pay structure offers a Pareto improvement: in addition to substantially increasing achievement, it improves teacher welfare by seven percent. Salary increases come at the expense of lower replacement rates in retirement and shifts toward defined-contributions plans which are preferred by teachers while being less costly to schools than are defined benefits plans.

The preferences of marginal teachers are especially relevant, and measuring their preference helps evaluate the generalizability of the estimates. Marginal teachers are not only the germane margin of labor supply, but also have higher academic ability and value-added. Thus, their retention decision affects the quality distribution of the teacher workforce (Weaver 1979; Schlechty and Vance 1981, 1982; Stinebrickner 2001; Wiswall 2013; Wheelan 2019). To explore the preferences of marginal teachers, (1) I test whether teachers who *eventually* leave the district

[9] Here, tenure refers to how long a person remains in teaching, equivalent to the average experience level.

have the same preferences as those who remain; and (2) I survey college students in the vicinity of 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, supporting the view that marginal teachers have the same preferences for compensation as those we measure but have a lower taste for teaching.

This study continues literatures that explore teacher preferences (Antos and Rosen 1975; Ballou 1996; Boyd et al. 2013; Biasi 2019), teacher compensation (Hanushek 1986; Card and Krueger 1992; Ballou and Podgursky 1995, 1997; Figlio 1997; Loeb and Page 2000; Hendricks 2014), teacher quality (Rockoff 2004; Hanushek and Rivkin 2006; Jackson 2009; Chetty, Rockoff, and Friedman 2014), and selection (Lazear 2000; Staiger and Rockoff 2010; Winters and Cowen 2013; Rothstein 2015; Baron 2020; Brown and Andrabi 2020). For a lack of options, previous studies have largely relied on either exploratory simulations, or used equilibrium data to estimate preferences. The first is a kind of principled conjecture. The latter suffers from a host of confounding factors. In this study, preferences are estimated using a field experiment and those preferences animate a model of teacher behavior grounded in real data. Moreover, prior studies are not able to estimate willingness-to-pay for most important components of teacher compensation and working conditions since they do not vary independently in the real world.

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 would be insufficient variation in naturally occurring records. By measuring preferences for a comprehensive set of attributes, I can calculate "preferred" structures and evaluate their effect. Mine is the first study to use choice data to calculate policy experiments for compensation structure and working conditions, and I argue it does so in a profession of pivotal importance. Finally, this paper demonstrates that compensation structure may be an effective tool for policymakers, not by eliciting effort, but by influencing selection.

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 simultaneous competing offers.[10] 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 difficult work settings (Antos and Rosen 1975). In either case, the resulting estimates would not be useful for predicting the effect of policy changes.

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 shared effect of union bargaining. Districts within a state often share a pension program, health-insurance plan, class-size regulations, and salary schedules. 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.[11]

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

[10] The job market is highly decentralized, so schools make offers at widely varying times; since offers explode within 24 hours, teachers rarely entertain two or more 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.

[11] 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.

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 to me. (5) The experiment is anonymous, so social-desirability bias is 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 conditions.

I expand on the main points.

First, whereas questions asking for open-ended willingness-to-pay 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;[12] 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 the need for no new effort to discover them. Several papers 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 an example, if teachers pay part of

[12] 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."

their health insurance premium, they should be indifferent between an additional dollar of salary or an additional dollar offsetting what they pay for insurance since the tax treatment is identical. Teachers value health-insurance subsidies *identically* to an equivalent increase in salary. Similarly, the discount rate that rationalizes teachers' salary-retirement tradeoff is *exactly* that estimated in the empirical literature on discounting (Best et al. 2018; Ericson and Laibson 2018). And the cost of commuting implied by my data matches a developed urban literature estimating the same.[13] The result holds when making within-teacher comparisons using teacher fixed effects. Therefore, in practice, the hypothetical choices teachers make demonstrate an amazing degree of realism.

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 or please an interviewer (Atkin and Chaffee 1972; Campbell 1981; Cotter, Cohen, and Coulter 1982; Finkel, Guterbock, and Borg 1991; Fisher 1993; Krosnick 1999). Surveys were 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 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 she chose 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 I partnered with; 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 their preferences in earnest (Carson, Groves, and Machina 2000; see also Vossler, Doyon, and Rondeau 2012). Carson

[13] Moreover, I test whether choice is monotonic in ordered variables that have clear impacts on utility (Hainmueller and Hiscox 2010). I find that choosing an offer is strongly increasing all along the support in starting salary, salary growth, retirement replacement rate, class-size reductions, and support provided to teachers.

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 among subjects (Bulte et al. 2005, Carson et al. 2006, Herriges et al. 2010).

Implementation

This paper aims to estimate 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. 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.[14] The choice experiment allows the analyst to evaluate several hypotheses in a single study and, importantly, compare the influence of various factors within a shared setting, making the estimates directly comparable.

In this survey, 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

[14] 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 essentially equally costly.

that allow us to assess the realism of responses. Attributes take on several values, shown in online Appendix table 1.[15]

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 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 1–3.

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 in the choice sets a placebo feature that should have no plausible impact on teacher utility—whether the school bus at the featured school

[15] 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.

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, similar to 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 the outcome variable—their choice—which does not introduce bias (Wooldridge 2010).[16] 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 variation in attention drives the result.[17]

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

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

[17] 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 ($p < 0.001$).

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.

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: $X_j = [X_{j1}, \ldots, X_{jK}]$. These offer attributes include compensation structure and non-pecuniary attributes. To explore the influence of each factor, I use conditional logistic models as well as linear-probability models to estimate utility, regressing 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:

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

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 in accurately predicting the choices of subjects. For example, the LPM accurately predicts 64 percent of choices in the main deck, whereas the conditional logit predicts slightly less, at 62 percent. 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 and its compensation. Aldine Independent School District instructs 70,000 students each year in Houston, Texas, with an annual budget of $700 million dollars (USDOE, 2016; NCES, 2019). Over three-quarters are eligible for free school meals (77 percent), which places them 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 hired by the district to deliver recommendations which, in 2016, included my experiment. The average teacher in the district has nine years of experience, and 30 percent of them have advanced degrees. Just over two-thirds are female (68 percent); the plurality is black (37 percent), and the remaining teachers are white (28 percent) and Hispanic (21 percent) (online Appendix table 2).

III. Teacher Preferences for Compensation and Other Factors

The first movement of the paper measures teacher preferences for each of several attributes of their compensation and working conditions. By estimating willingness-to-pay, we can later assess whether teacher welfare can be improved by reallocating compensation structure.

The main preference estimates are presented in figures 1–3 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 and compare the influence of different offer characteristics. In table 2, I use the continuous variables and present part-worth utility βs and translate them to an interpretable willingness-to-pay (WTP) for each trait; the left three columns represent estimates from a linear probability model, and the right three represent estimates from the conditional logistic model estimated with maximum likelihood. All estimates are standardized across the three decks using subjects' responses to the salary feature.[18] 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 if she remains six years does the total

[18] Specifically, each coefficient in Deck 2, for instance, 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 $\beta_{salary}^{Deck1}/\beta_{salary}^{Deck3}$.

present value of an additional 1 percent growth exceed the total present value of $1,000 higher in starting salary.[19]

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 secure 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. Reassuringly, 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).[20] This reinforces the claim that teachers respond realistically to the experiment.

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 its cost. 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 (VAM) 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 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 also test whether teachers' aversion to rewards based only on VAM differs by whether the teacher has a relatively low VAM compared to their Danielson score. Preferences do not differ

[19] Interestingly, teachers in the district have on average just over six years of experience, again suggesting the realism of teacher responses.
[20] The WTP for retirement income by new teachers is slightly lower, but implies a similar δ of 0.939.

by relative strength on VAM or Danielson, suggesting that teachers preferences for evaluation form are not strategic.

The presented job offers vary how many years teachers are evaluated before granting 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 equivalent to a $470 salary increase. 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 discernible 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 be paid $9 per hour to commute to work, half a teacher's hourly wage ($19)—exactly consistent with the literature on the cost of commuting (Small 2012; Mas and Pallais 2017).

Reducing class size by one student increases 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, though these estimates do not rely on experimental designs. 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 for class size. 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 preference is possibly related to the costly nature of transferring tasks (Goldin 2014). The WTP for the first few hours of help is higher than the average WTP, suggesting that providing a low level of assistance would be cost effective. The third deck varies student attributes and school-leadership characteristics. Teachers prefer schools with higher-achieving students and fewer children in poverty, but they have no preference over the racial composition of their students, consistent with Antos and Rosen (1975) who find the same pattern.

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 contact a random sample of respondents, who indicate 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 costs of working in a low-achieving school and 85 percent of the disutility associated with teaching in a high-poverty setting (table 3). This suggests both that disruptive students are perceived by teachers as costly and that principal support is highly effective in mitigating those costs.

IV. Using Compensation to Affect Selection

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

Whether or not compensation and working conditions can generate a “separating equilibrium” in which high-type teachers differentially select into, or differential remain in, a school depends on whether excellent teachers have *distinctive* preferences. Perhaps high-quality teachers have weaker aversion to long probationary periods (worrying less about dismissal), stronger preferences for small classes (placing a higher value on individual attention), greater value on high starting salaries (having stronger outside options), or distinctive appreciation for generous pensions (being more committed to a long career in teaching) (Morrissey 2017; Weller 2017). It’s also important to know whether highly rated teachers have different preferences for working conditions that are not affected by policy—such as student demographics—to understand whether larger compensating differentials are needed to draw high-performing teachers to low-income schools.

To measure teacher performance, I estimate value-added models (VAM) 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 VAMs using all the available test scores that a student has from their previous school year while controlling for student fixed effects, school-year fixed effects, and indicators for whether last year’s test score is missing in each subject. The VAM used in the primary analysis is the average of the subject-specific VAMs available, usually math and reading. The resulting VAM measure is 0 on average with a standard deviation of 1. I sort teachers into ten deciles based on their VAM and generate a quality index from those deciles from 0 to 1. Since students are not tested in all grades, there are records to estimate value-added for just under half of teachers. To provide a measure of quality that covers a broader array of teachers, I use 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 VAM index and the Danielson index are significantly correlated for teachers with both measures ($p < 0.001$). For those teachers who do not have a VAM index, I input the Danielson index as their quality measure. Together, the VAM index and the Danielson index provide a quality measure for just under 80 percent of respondents. I find the same results when using either measure in isolation.[21]

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).

The most highly rated teachers have similar preferences to their colleagues for most school attributes (table 4 and online Appendix tables 6 and 7). 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 merit rewards (table 4 and figure 4). 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 $p < 0.001$).[22] If teachers entertained two comparable offers, a high-

[21] This finding also holds when using only VAM or only Danielson observation scores, shown in online Appendix table 8.

[22] 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,

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 the presence of performance pay, at least in retention. Since the best performers receive increased compensation, the probability of attrition is reduced relative to teachers with lower performance. Whether merit rewards can generate favorable selection *into* teaching is not clear from this study. Performance pay may not affect selection on entry if prospective teachers do not know their ability to teach.

The relationship between teacher quality and preferences for performance pay is illustrated in figure 4. 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. I present the corollary plot for each of the other attributes in online Appendix figures 4–6, each of which lack a systematic pattern, findings that are consistent with the results in table 4 and in online Appendix tables 6 and 7 in which higher quality teachers do not differ significantly in their preferences for school attributes. In future work, it may be fruitful to study whether there are differential preferences for other attributes including dismissal rules and measures of colleague quality.

The Preferences of Marginal Teachers

An important dimension of heterogeneity is whether marginal teachers (those close to indifference between remaining in the profession and exiting) 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 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 systematically different preferences in two of those attributes, both significant at the five-percent level. Leavers have slightly weaker aversion to large classes and

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.

slightly stronger interest in having teaching aids. In all other attributes, leavers have statistically identical preferences (online Appendix tables 18–20).[23]

To explore whether the preferences of marginal teachers differ on entry, I survey 1,193 college students in a large public university near of the district. Students are asked to describe how likely they are to teach (on a Likert scale from "I would never consider teaching" to "I've never considered it, but I'd be open to it" to "I've thought about teaching" to "I've considered it seriously" to "I plan to be a teacher"). I ask the respondents to imagine that, regardless of their interest in teaching, they decided to become a teacher at least 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 for our purposes *here* is whether those planning on teaching have similar preferences to marginal teachers—those considering or open to it. Preferences are similar throughout the spectrum of interest in teaching. 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 gamut of interest in teaching, preferences differ little along the teacher-propensity index. The joint significance, for instance, of attributes interacted with the teacher-propensity index is jointly insignificant in the main deck. Areas in which inframarginal teachers seem to differ from other respondents tend to be in attributes on which 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 on teaching. This exercise implies that tastes for compensation structure are largely uniform along the distribution of interest in teaching, suggesting that the preferences uncovered in the experiment generalize to *marginal* teachers on the entry and exit margins. What differs is their tastes for teaching, not their tastes for compensation.

V. Optimizing the Compensation Structure

In this final movement of the paper, I use the estimates I've developed describing teacher preferences and selection to calculate the policies that would maximize various objectives, principally maximizing student achievement.

[23] I also test whether preferences differ by grade level. In general, teachers in elementary schools, middle schools, and high schools have similar preferences for compensation, student attributes, principal affect, commuting, and assistance. Middle and high school teachers, however, express less aversion to large classes and stronger aversion to longer tenuring periods than elementary-school teachers (online Appendix tables 21–23).

Compensation Structure

What do preferences suggest about how schools should structure compensation? 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 affecting retention to maximize teacher tenure. Third, I calibrate a model of the achievement production function from the literature that includes 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 influences achievement by affecting the effort of teachers (Lavy 2002, 2009; Imberman and Lovenheim 2015; Biasi 2019) and by differentially retaining high-performing teachers (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 performance pay.

All the 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 implicitly assume that incoming teachers have similar preferences and ignore the effect of compensation structures on selection on entry. Since preferences are the same along the propensity-to-teach index, this assumption is accurate. If anything, the assumption understates the influence of a compensation structure on achievement if performance pay induces 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 24 and 25).

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

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 the *current* faculty, which appears to be a realistic goal for union schools since their operations tend to serve incumbent teachers (Hoxby 1996; Figlio 2002).

Maximizing tenure (the average experience of teachers) is a related objective, and teacher experience is one of few reliable predictors of teacher performance. 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 plug into my utility calculations which are also in units of salary. Importantly, because the estimates in Hendricks (2014) come from Texas, they immediately 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 changes with experience. The retention probability at experience level *e* is calculated:

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

The Δ_a is the difference in utility between the compensation in Hendricks and the salary-equivalent utility of the bundle under consideration, U_a from equation (2), where the difference enters the model as a percent. To help the reliability of these comparisons, I scale the salary schedule prevailing in Hendricks by an additive term so that the average tenure predicted from the model using status quo compensation matches the district's actual average experience (9.03 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 $\mathbf{D}_e = [D_1, D_2, \ldots, D_{35}]$ where D_e states the share of teachers with experience level *e*. 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 central objective function I consider is the maximization of student achievement. In the achievement 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 having increasing, concave impacts on students (Papay and Kraft 2015). To simulate the influence of experience on achievement, I calculate retention probabilities, as above, and then simulate the equilibrium experience profile and take the dot product with VAM-over-the-life-cycle vector from Papay and Kraft. Performance pay improves achievement somewhat 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, which is not a feature we observe in the real world. After calculating the quantity who exit each year from the retention probabilities r_{ea}, I remove teachers with the lowest calculated 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 $\mathbf{Q}_d$, 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 that 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 better targets utility 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 in Houston. Performance pay increases achievement through selection 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 achievement rises by 0.023 standard deviation from selection for an additional $1,000 in performance pay. The effect of teacher retention is merely the dot product of the experience distribution and experience-specific value-added measures.

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

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

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, which is calculated by summing the departures calculated when simulating the experience distribution (Barnes, Crowe, and Shaefer 2007; Watlington et al. 2010). Retention is partly budget saving. Some small per-person costs, *U*, are required, which captures the cost of unemployment insurance and workerman'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 principle 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 (with which to divide students among teachers) become more expensive.

I also constrain the optimization exercise 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. 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-VAM-only evaluations) are constrained to be within $[0,1]$. I go into detail about the maximization exercise in online Appendix D.[24]

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 displayed in Table 5).

Compensation Structure to Maximize Teacher Utility

At the time of the survey, the district paid $50,000 in base salary, with a 1.8 percent average yearly increase in salary earnings. They provided no performance pay, had an average class size of 28.7 students, paid $3,960 in health-insurance subsidies, and 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

[24] 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.

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 on other outcomes 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 0.066σ each year, which comes from increased teacher experience (31%), induced effort from merit pay (30%), and increased retention of high-caliber teachers (38%).

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 teachers improve throughout their career (Wiswall 2013; Papay and Kraft 2015). Districts could structure compensation and working conditions to promote long tenures of their teachers.

The compensation structure that maximizes (average) experience 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. When I require the district to use a pension, the solution replaces 25.5 percent of salary in retirement instead of 56.6 percent. These lower replacement rates overstate the reduction in retirement income since the replacement rate applies to a higher final salary.[25] 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

[25] 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.

influence of pension wealth accumulation on attrition. Pensions benefits are backloaded, so they produce a strong pull for teachers nearing eligibility, when pension benefits spike, but they do little to retain younger teachers. These simulations suggest that defined contributions plans, on net, *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 the main mass of teachers in DC plans than in pensions.

The "optimal" structure for maximizing average tenure 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 0.067σ increase in student achievement each year, an improvement that arises from an increase in teacher experience (32%), an increase in teacher effort from performance pay (30%), and positive selection in retention (38%).

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 rate (1.3 percent growth rate), $5,000 available each year in performance pay, and a class size that's 3.5 percent larger. Whereas the other optimizations recommended using VAM in combination with observation scores to distribute performance payments, this model recommends using VAM-only to evaluate performance. This practice improves targeting payments to high-VA teachers to reduce their attrition.[26] The 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 by 0.194σ per year while also improving teacher welfare by 7 percent at the same time. The achievement gains come from more experienced teachers (5%), effort induced by merit pay (35%), and improved retention of high-caliber teachers (60%).

These reforms are simulated based on a partial-equilibrium framework in which one district adopts the estimated structure that is assumed to have no impact on the selection *into* the school

[26] When performance pay influences selection (on entry or exit), the standard for being in the top quarter evolves. Schools could fix the standard by benchmarking VA measures to the distribution of VA in districts that do not implement VA, or they could benchmark VA so that scores that would have qualified as being in the top quarter of the original distribution are still rewarded.

district. This leads to a suitably conservative estimate. The achievement gains are fully realized in time by affecting retention patterns. One question of interest is whether performance pay can generate positive selection into teaching if broadly adopted. 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, those open to teaching, must have similar (affirmative) preferences for performance pay as other teachers. Both in the district and among prospective teachers, I find that marginal teachers have identical (affirmative) 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).

As would be true in a survey of any district, the experimental variation reveals the preferences for a given group of workers, possibly because of the compensation structure already in place. 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 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 so clearly among an endogenously selected group implies 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. Although it has weak union presence, it may be that political constraints or bargaining distort compensation. 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).[27] If true, we might expect places with stronger union presence to pay a larger share of compensation in benefits, conditional on total compensation.[28] 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 several 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 ($p < 0.001$), 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 26).

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.[29] One of the consistent lessons from the maximization exercise is that the district can improve teacher welfare, tenure, 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. In online Appendix figure 9, I show where the district's compensation falls in the distribution of US districts in terms of salary share. Two-thirds of school districts have salary shares below the district; 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 less are likely also underinvesting.

The results highlight several avenues for future work. Work that examines *entry* and *exit* in the profession would provide an 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 larger numbers of highly able young people into the profession. To examine exit,

[27] 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.

[28] There is a strong negative relationship between total compensation and salary share, perhaps since other amenities become more important as the value of a marginal increase in salary diminishes. 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.

[29] 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. See, for reference, online Appendix figure 10.

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 study whether excellent teachers have distinct preferences for colleague quality, dismissal risk, or other attributes may provide policymakers with additional tools to recruit and retain excellent instructors. Research to evaluate whether the preferences we report here are comparable to teacher preferences in other areas of the country would be useful for discerning how general these preferences—and their implications—are. 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 (independently) 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 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, 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 meaningfully alters the selection in over time (Lazear 2000). Other compensation, contract, and working-condition attributes provide little scope for enhancing selection.

Using the model of teacher utility and the scope for shaping 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 closer to one another than any are to current structure. Each implies a higher allocation of 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 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).

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.

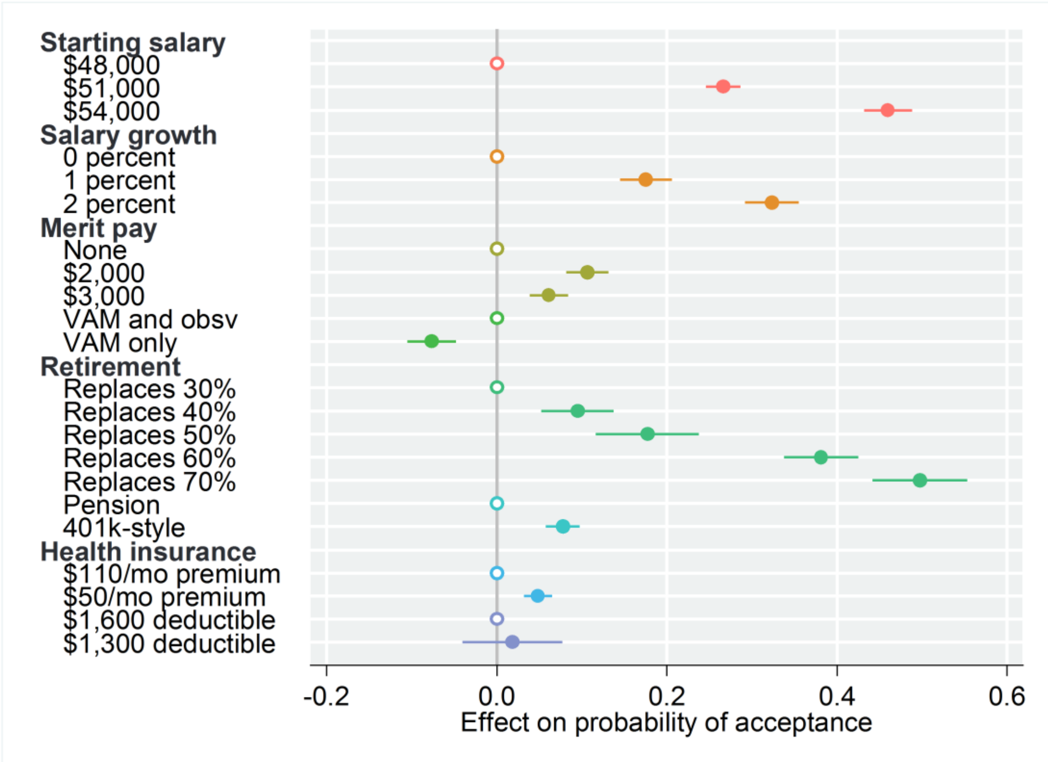
Winters, Marcus A., and Joshua M. Cowen. “Would a Value-added System of Retention Improve the Distribution of Teacher Quality? A Simulation of Alternative Policies." *Journal of Policy Analysis and Management* 32, no. 3 (2013): 634-654.

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.

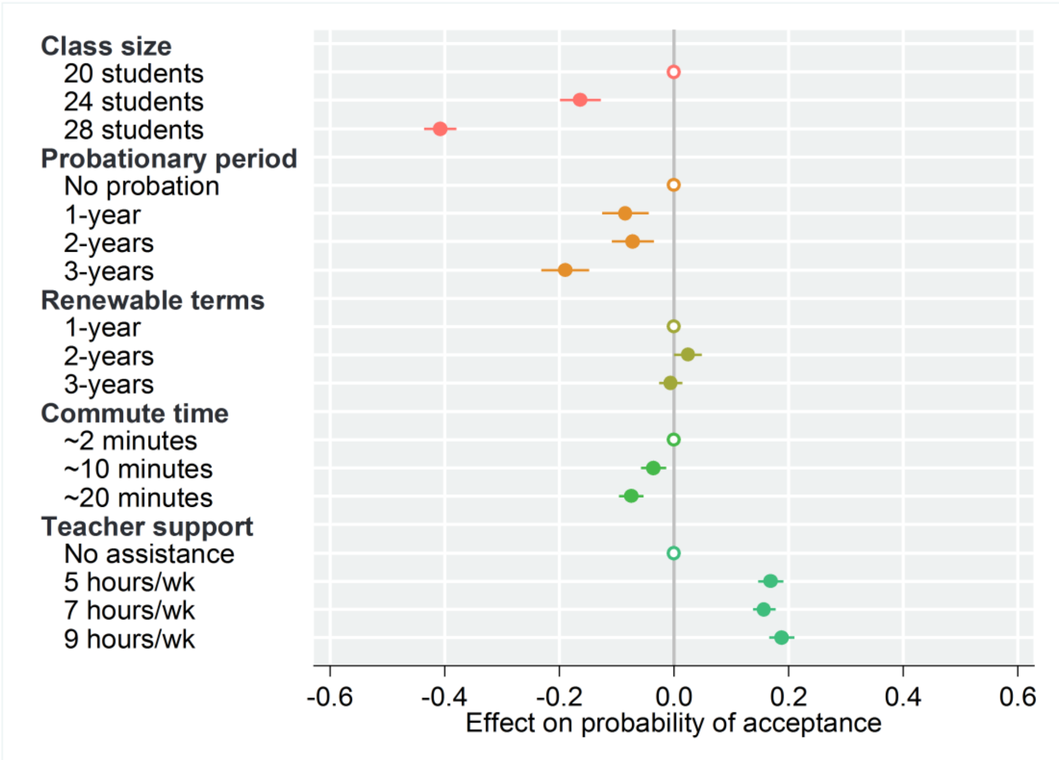
Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. MIT press, 2010.

FIGURE 1—EFFECTS OF COMPENSATION ATTRIBUTES ON THE PROBABILITY THAT TEACHERS ACCEPT A JOB OFFER



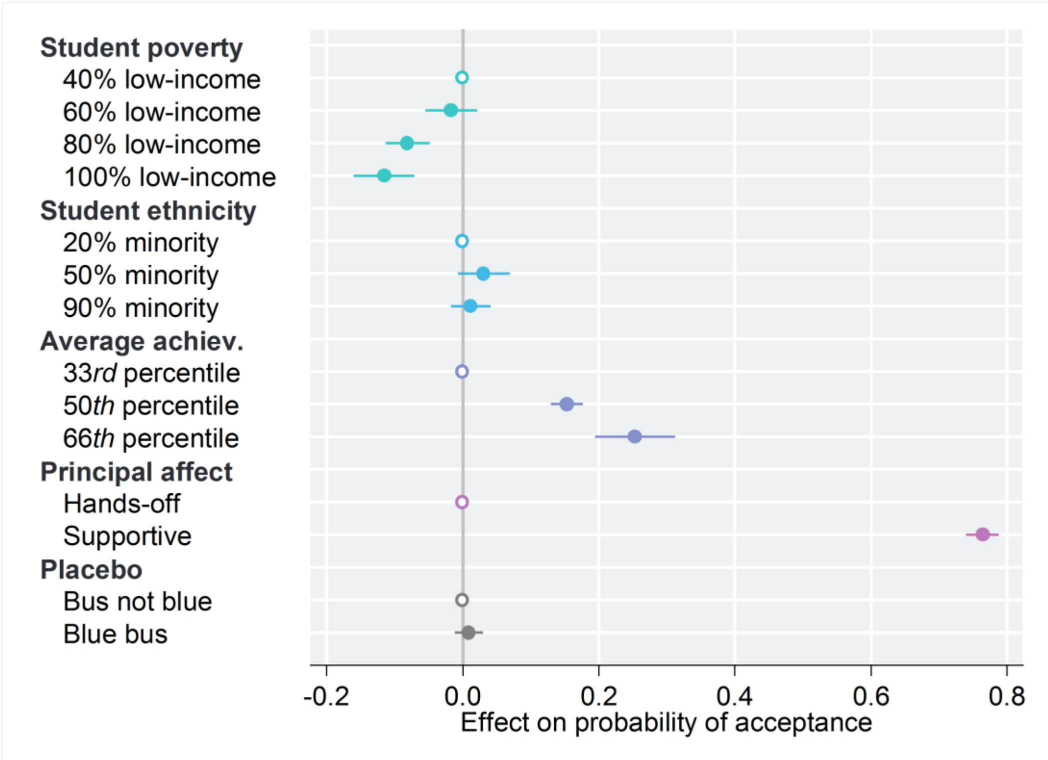
Note: Dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from least-squares regression. The unfilled dots on the zero line denote the reference category for each job-offer attribute. Online Appendix Table 2 displays the underlying regression results.

FIGURE 2—EFFECTS OF WORKING-CONDITION ATTRIBUTES ON THE PROBABILITY THAT TEACHERS ACCEPT A JOB OFFER



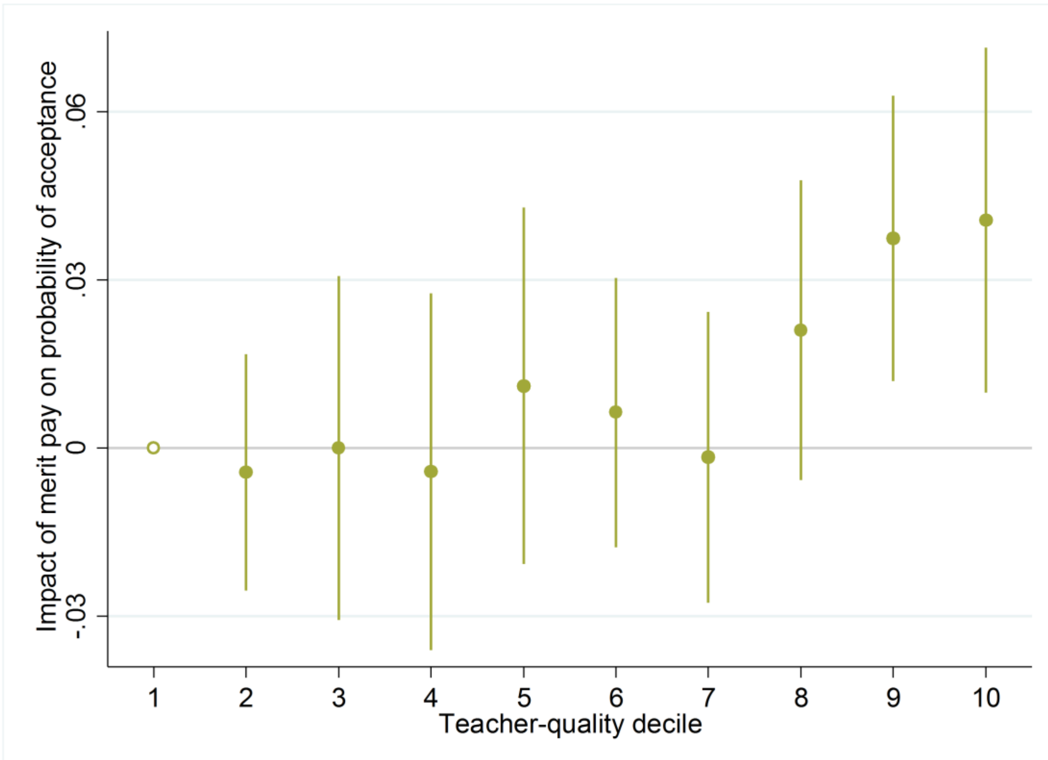
Note: Dots with horizontal lines indicate point estimates with cluster-robust, 95%-confidence intervals (CI) from least-squares regression. The unfilled dots on the zero line denote the reference category for each job-offer attribute. Online Appendix Table 3 displays the underlying regression results.

FIGURE 3—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. Online Appendix Table 4 displays the underlying regression results.

FIGURE 4—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 VAM and, for those teachers who lack a VAM 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
VAM 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

	Linear Probability			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
VAM 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?

	LPM (1)	LPM (2)	LPM (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	Quality-index interaction	Reference Group	Quality-index interaction
	(1)	(2)	(3)	(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)
VAM 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**
VAM-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	0.158σ	0.158σ	**0.286σ**

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.

ONLINE APPENDICES

Preferences, Selection, and the Structure of Teacher Compensation

Andrew C. Johnston

Online Appendix A: Estimation of Value-Added Measures

In the empirical analysis on separating equilibria, we divide teachers into bins based on their value-added measure (VAM). In this online Appendix, I discuss the methodology for estimating VAM for teachers in Aldine ISD.

The school district provided student-teacher linked test score records from the 2011–12 school year through to the 2015–16 school year, covering some 60,501 students and 3,559 teachers. These files contain yearly student performance on the STAAR exam (State of Texas Assessments of Academic Readiness) administered statewide by the Texas Education Agency. STAAR tests mathematics, reading, writing, science, and social studies, depending on the year. The state tests reading and mathematics in grades 3–8; writing in grades 4 and 7; science in grades 5 and 8; and social studies in grade 8. Like commonly used VA models, I estimate teacher value-added using the equation

$$A_{istm} = f\left(A_{i,t-1}\right) + \delta_{st} + \alpha_i + \gamma_m + \varepsilon_{istm}$$

I parameterize the control function for lagged test scores using a linear expression of prior-year scores in all available subjects, with indicators for whether the student lacks scores in each subject. To account for student-specific student achievement trajectories, I include student fixed effects, α_i; and control for school-year differences in achievement gains with school-year specific fixed effects, δ_{st}, to capture yearly school/neighborhood effects that are unrelated to the teacher assignment. The parameters γ_m capture average teacher-specific contributions to student achievement, holding all else constant, which I take as the measure of teacher value-added.

Online Appendix B: Cost Function of Compensation Structure

Crucial to calculating the optimal structure of compensation and working conditions is properly specifying the cost as a function of each element. In this Appendix, I provide detail on how the cost function is constructed.

Salary

Because Aldine ISD does not participate in Social Security, they pay modest payroll taxes. Both in documents from the district and in the district's financial disclosures, the district pays 1.5 percent of its payroll in payroll taxes, approximately the rate owed for Medicare taxes, 1.45 percent. Thus, the cost of an additional $1 in salary compensation costs the district $1.015. The cost of salary provision also interacts with the cost of salary growth and retirement, discussed below.

Health Insurance

In July 2016, three months after the survey was administered, I collected data from the Affordable Care Act (ACA) health exchange which indicated the monthly premium, deductible, cost of an office visit, and plan type (HMO, PPO, POS, PD, catastrophic) for 50 plans available in the Houston area. A hedonic pricing model revealed that the cost of office visits (the copay) had no systematic relationship with price (premia), which was most predicted by the deductible ($p < 0.001$) and HMO status ($p < 0.001$). With no deductible, a generic plan cost \$385.70 (CI: \$361.34 – \$410.06) per month, and the cost declined by \$24.40 (\$20.30 – \$28.49) for every \$1,000 increase in the deductible. There is no evidence that the price is a quadratic function of the deductible.[30]

$$\text{Annual Cost} = 12 \times (385.7 - 24.4 \times \text{deductible})$$

In my model, I use the value of insurance subsidies, in part because we do not have enough power or variation to precisely pick out the "right" health plan. Moreover, in practice, teachers have an insignificant preference in favor of dollars paid in salary over dollars paid in health insurance, meaning that, when optimizing teacher utility, the school district shifts away from health insurance compensation, allowing teachers to privately optimize their insurance decision.

Merit Pay

The merit compensation teachers are offered in the survey is paid to "the top 25 percent of each school based on principal ratings and student growth." Because performance compensation is paid only to a quarter of teachers, the cost of providing an additional \$1 in merit pay is \$0.25 per teacher. This income is subject to Medicare taxes, 1.45 percent.

Defined Benefits Plan (Pension)

The explicit promise of a *defined benefits* program is that it is not subject to market risk—the benefit is guaranteed, rather than the contribution, is fixed. Marx and Rauh (2014) show that, in order to satisfy the funding requirements, pension managers assume a constant, high rate of growth (7.5–8.0 percent) with no risk in order to balance their revenues with expected demands. This leads to underfunding above and beyond the shortfall recognized under even these optimistic assumptions. The actual return of an essentially risk-free investment, like treasury bonds, is 1.57 percent in normal times (now much lower). I assume a rate of 1.57 percent and calculate what would be saved by retirement's onset if a teacher were setting aside 1 percent of her wages each year. I then take the lump sum accumulated by retirement (assumed at age 65) and annuitize it, using an annuity calculator.[31] I then take the annual annuity as a fraction of the teacher's highest

[30] When the quadratic term is included, the coefficient's p-value is 0.688.
[31] http://money.cnn.com/tools/annuities/

salary to make a mapping from what percent of salary the teacher is saving to her replacement rate. With a 1.57 percent risk-free rate of return, a one-percent saving pattern replaces two percent of the teacher's salary, meaning that teachers must save 0.559 percent of their income to finance an additional percentage point of replacement rate under a risk-free rate of return. Pensions, however, enjoy a cost saving since some teachers will pay into the pension but will not persist long enough to vest and receive an annuity. I calculate the share of those paying into the pension each year who will leave *before* the vesting period is complete. That fraction is then applied as a discount on the cost of the pension.

Defined Contributions Plan (403(b))

Nonprofit and governmental agencies can provide a retirement plan that is corollary to the 401(k), called the 403(b), which are available to all tax-exempt organizations. In 403(b) accounts, the school commits to contributing a defined amount to the worker's retirement rather than promising a defined level of benefits at retirement. While pensions take several years for a worker to vest and retirement benefits are heavily backloaded,[32] 403(b) plans accumulate retirement wealth proportional to employment and vest immediately, making retirement contribution totally portable. I follow the same calculation as described above to generate the cost of an average replacement rate through the 403(b), but use as the expected interest rate 7.5 percent, under the historical trend (ten percent) (Cowen 2011; Gordon 2016). Here from, the cost of saving enough to replace one percent of a teacher's salary (in expectation) is 0.220 percent of your salary. If one assumed an eight-percent return, the coefficient on rep would be 0.202 percent.

Class Size

One of the chief conceptual issues in structuring the cost function is how to formalize the cost of class-size choices while allowing compensation to vary flexibly. For instance, by simply using the average cost of class-size reductions from a paper, the analysis would not account for the fact that class-size changes become more and less costly based on the costliness of the compensation package itself. The fundamental problem is that reducing class size requires hiring an additional teacher, the cost of which depends on the cost of the compensation package. Moreover, the cost of additional class-size reductions increase quadratically as class size falls. To accommodate this tradeoff in optimization, I conceptualize the cost function as a joint choice of compensation structure (which determines the average cost per teacher) and class size (which determines the

[32] Vesting refers to when the employee becomes eligible for retirement payments even should they retire or quit. The granting to an employee of credits toward a pension even if separated from the job before retirement.

number of teachers needed), allowing the cost structure of teacher pay to flexibly affect the cost of class-size adjustments. To provide a smooth function for optimizing, we model teacher quantity as continuous.

Endogenous Retention

What makes the calculation of the cost of salary growth rates somewhat complicated is that providing more generous compensation reduces attrition, increasing the cost both through salaries *and* by increasing the odds that teachers are retained to be paid at higher steps of the salary schedule. Hendricks (2014) estimates the effect of additional salary on the attrition probability of teachers at different points of their experience profile and finds that compensation has significant impacts on attrition for new teachers which influence declines as teachers approach veteran status. His study uses data from Texas, and it's fortunate to have estimates on the impact of compensation on retention, throughout the teacher life cycle, from the labor market in question.

To adjust for the cost of endogenous retention, I calculate the total utility of teachers with status-quo compensation and difference it from candidate compensation structures. I multiply those differences by turnover elasticities for teachers of every experience level from Hendricks, which generates a vector describing how the new compensation structure would affect turnover at each experience point. I add these adjustments to the natural turnover rate and then calculate the steady-state distribution of teacher experience based on the affected retention patters. This allows me to construct the average compensation cost in steady state, a function of compensation and the (endogenous) distribution of teacher experience.

Cost of Turnover

A related element affecting the cost of lower retention and reduced class size is the fixed costs of employing an additional teacher, the primary of which is more frequent hiring and training. Barnes, Crowe, and Shaefer (2007) and Watlington et al. (2010) study the costs of turnover in schools in terms of recruiting, screening, and training. The authors do an in-depth accounting exercise with five school districts and find that a typical new hire costs \$11,891, on average, in screening, processing, and onboarding. Because the average teacher turns over every 6.13 years (the average years of experience in Hendricks (2014)), the yearly cost of hiring is \$1,938 per teacher each year under the status quo retention pattern. I allow retention patterns to evolve in response to compensation and working conditions and explicitly calculate the cost of turnover based on the share of teachers that attrit in a year multiplied by the number of teachers times the cost of replacing each.

I calculate other fixed costs of employment, but they are trivial. The wage base of unemployment insurance is smaller than the typical yearly salary, so UI taxes function effectively as a head tax of only $11 per teacher per year in this district (calculated from financial disclosures). The district also pays $167 per teacher per year for workers compensation. A final consideration is the costs for space. Throughout, I use as the benchmark a sort of steady state. If a class is made smaller, I assume that each classroom can be made smaller costlessly, either in new construction or in a one-time construction cost. It may be that teachers have their own office space in some districts, but I ignore this cost for simplicity.

Online Appendix C: Preference Heterogeneity

Here I explore how preferences vary by teacher race, sex, and experience. A considerable body of work finds that students progress more quickly when taught by experienced teachers and those whose race or sex matches their own (Dee 2004, 2007; Bettinger and Long 2005; Clotfelter et al. 2006; Carrell et al. 2010; Kofoed and McGovney 2017; and, in particular, Gershenson et al. 2018). Understanding how preferences vary by group may help schools attract and retain a desired demogrphic.

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

More experienced teachers have weaker preferences for higher salary and stronger preferences for more generous retirement plans (online Appendix table 9). In working conditions, preferences are similar to those of novices in time-to-tenure, term length, and commute time, but older teachers have a higher tolerance for larger classes and a stronger demand for teaching assistance. Senior teachers also have stronger preferences in favor of high-achieving students than their less experienced colleagues. Novice, new, and experienced teachers have similar preferences for having a "supportive" principal, but veteran teachers place an additional premium on it (online Appendix tables 9–11). In principle, a district could attempt to retain veteran teachers by providing compensation options that suited the preferences of established teachers.

Black-white and male-female achievement gaps may partly be the byproduct of skewed teacher demographics (Goldhaber et al. 2019). Understanding how preferences differ by group may help districts attract and retain teachers of a particular group (for instance, to retain experienced teachers or to tilt the sex (race) distribution of teachers to mirror the distribution of students).

I follow the same course to study how preferences differ by sex, interacting male indicators with each attribute. Men have stronger preferences for salary than women and are more averse to high-deductible health plans, consistent with women being more likely to receive health insurance through a spouse. Like senior teachers, men are more willing to teach large classes, but they place a lower value on assistance with grading. Men and women have similar preferences for student demographic characteristics, but men exhibit less demand for a supportive principal (online Appendix tables 12–14). I also explore how preferences differ by race. Black teachers have weaker preferences for salary growth than white and Hispanic teachers. Black and Hispanic teachers have stronger preferences for performance pay than white teachers. Black teachers place higher value on a short tenure clock and less frequent reviews than white and Hispanic teachers. All three groups have similar preferences for commuting and assistance with grading. While white and Hispanic teachers have precisely zero preference for student race, black teachers prefer student bodies that have a higher minority share, mirroring Antos and Rosen (1975). While everyone has strong preferences for a supportive principal, black and Hispanic teachers value supportive principals 8–12 percent less than white teachers (online Appendix tables 15–17). That both male and minority teachers have weaker preferences for principal support suggests they either experience lower costs of classroom disruption or enjoy additional social capital with disruptive students.

Online Appendix D: Objective Functions

Teacher Utility

As a kind of baseline, I use as the objective function the teacher-utility model estimated from the data, essentially acting as if the district's goal is to structure conditions to maximize the wellbeing of teachers, subject to the budget constraint. This may also be similar to the stated goals of a teachers' union. This model provides some of the core influence of the other optimization criteria because teacher utility affects the retention probabilities that influence, for instance, achievement. I estimate the model of teacher utility (the coefficients from simply regressing teacher choices on attributes) with nonlinearities for merit pay, growth rate, replacement rate, and class size; these nonlinearities prevent compensation from loading into the attribute with the highest average return.

When the maximization is unfettered, class size balloons to pay for higher salaries. In Texas, classes can be no more than 22 students for students from kindergarten through fourth grade, but there is no statutory requirement for more advanced students, though legislation was proposed to limit class sizes to no more than 28 students for students in fifth through eighth grade (Green 2014). While the structure of other elements of compensation have little direct impact on students, class-size reductions are not intended, primarily, to appeal to teachers. For this exercise and those that follow, I limit the permissible range of class size to no more than 30 so that, should class-size reductions be an appealing improvement to teaching conditions, we can see those materialize in smaller class size, but not allow classes to explode in order to provide more generous compensation to incumbent teachers.

Teacher Retention

When teachers leave Aldine ISD, either by retirement from the profession or by transferring to another district, it opens a vacancy chain that results in the departed being replaced by a novice somewhere, which is quite costly to student achievement (Wiswall 2013). One objective that districts could pursue would be to structure compensation and working conditions to improve retention. I use the same basic structure used above to adjust for endogenous retention: retention probabilities are adjusted off a baseline based on how much the structure improves teacher utility. Using those adjusted retention probabilities, I simulate the share of teachers who will be in each experience cell in steady state. The dot product of experience shares and experience produces the average experience level (or tenure) with that structure of compensation, which is the object I maximize.

Student Achievement Production Function

What structure of pay maximizes student achievement rather than teacher satisfaction or tenure? I construct the achievement function to reflect the representative estimates of quasi-experimental domestic studies in terms of experience, class size, merit pay, and selection. I assume student achievement is a function of parent and teacher inputs, $A = g(P, T)$, where P reflects the input of parent and T reflects inputs of the teacher. The parents' impact, $P = h(t, r, k)$, is a function of the time parents allot to children (t), the resources made available to children (r), and the number of children the parents care for (k) (Price 2008; Loken, Mogstad, and Wiswall 2012; Black, Devereux, and Salvanes 2005). The teacher's role in achievement is a function of her innate teaching ability ψ, her skill σ which is influenced by experience ϵ and training τ, her effort e, and the size of her class c.

$$T = f(\psi, \sigma(\epsilon, \tau), e, c)$$

The teacher's skill increases quickly in experience ϵ before slowing its incline after the first few years. Traditional training programs have demonstrated little effect on teacher skill, though we might consider professional evaluations and mentoring programs a new generation of training (Taylor and Tyler 2012). Finally, effort is conceived as induced, unnatural effort—the increase prompted by incentive or accountability (Fryer et al. 2012; Imberman and Lovenheim 2015; Macartney 2016). In part because of limits in the literature, the achievement function I calibrate is a linearization in most arguments.

Experience

Retention affects teacher quality through two channels. First, teachers improve as they gain experience, especially at the beginning of their careers. If a given teacher turns over, the students she would have had will instead be taught by a novice who is systematically less effective. Second, early in the career, teachers with the largest positive impacts on students are the most likely to leave the profession. Thus, when increasing the retention odds, the stock of teacher quality improves both in experience and in composition because the marginal teacher to leave is, on average, of higher quality. In the basic model, we focus on the influence of additional experience improving a teacher's ability, since the effects of retention on the distribution of initial quality is somewhat unclear (Wiswall 2013; Hendricks 2018).

To quantify the influence of experience in the model, I rely on estimates from the discontinuous career model in Table 2 of Papay and Kraft (2015). I normalize average new-teacher VAM to zero and infer the typical teacher improvements in math and English (at five years, a typical teacher has improved 0.1216 in math and 0.0824 in English; by year 15, the typical teacher has improved an additional 0.1315 in math (suggesting that the typical teacher is 0.2531 better than a new teacher after having earned that much experience) and an additional 0.0831 in English (suggesting that the typical teacher with that experience is 0.1655 better than a new teacher)). Finally, the estimates suggest that teachers with 25 years of experience have improved from their 5-year experience level by an additional 0.2413 in mathematics and 0.1513 in English (0.3629 cumulatively in math and 0.1845 cumulatively in English by year 25).

To provide a general profile of experience on quality, I average the math and English returns. I fit a regression model of average VAM on experience and experience-squared using the first three experience nodes (0, 5, and 15), and a second model using the latter three points (5, 15, and 25) and use the predicted values (y-hat) from 0 through 5 in the first model and between 6 and 30 in the second model. Without the combination of these two piecewise models, the resulting experience profile either suggests convex increases in quality among veteran teachers, something never found in empirical work, or declines in quality among veteran teachers, which would contradict the estimates used to train the VAM profile in experience. The value-added profile that results from this procedure is most steeply increasing for new teachers but reflects the gains of experience throughout the life cycle of a teacher (Wiswall 2013; Papay and Kraft 2015). The resulting quality profile is presented in online Appendix figure 11.

Class Size

Analysts typically conclude that large class sizes reduce student achievement, especially for students that are young or low-income (Angrist and Lavy 1999; Krueger and Whitmore 2001; Jepsen and Rivkin 2009; Fredriksson, Ockert, Oosterbeek 2012, 2016; Schanzenbach 2014), but the literature contains a split (Hoxby 2000; Chingos 2013; Angrist, Lavy, Leder-Luis, and Shany 2019). In this paper, I incorporate domestic estimates of the influence of class size into the education production function. Krueger (1999) finds that an eight-student reduction (from 23 students to 15) increased achievement by 0.035σ per year, with larger effects in kindergarten (0.120σ), using random assignment from the Tennessee STAR experiment.[33] In contrast, Hoxby (2000) exploits natural variation arising from cohort sizes and class-size rules and finds no impact of class size on student achievement; her use of test scores after summer break may reflect rapid fadeout for class-size induced achievement gains. Dee and West (2011) use a within-student comparison for middle-school students and, similarly, find no overall impact of class size on student achievement. Cho, Glewwe, and Whitler (2012) follow Hoxby using recent data and find that a ten-student reduction in class size increases achievement by 0.04–0.05σ for students in elementary school, essentially in line with Krueger (1999). The domestic evidence tends to suggest class size does not matter as much for older grades and matters most for very young children. I take the average of these four estimates to predict that student achievement rises by 0.022σ for elementary students, with no effect of class sizes for students in middle or high school (Rivkin, Hanushek, and Kain 2005; Dee and West 2011; Chingos 2012). I use data from the National Center for Education Statistics to know what proportion of the district in question is a part of each school-

[33] The experimental setting may alter teachers' incentives, since the results of a known experiment may influence future working conditions.

type. The district serves a student body of 15.2 percent pre-school aged children, 37.6 percent elementary-school aged children, 22.5 percent middle school aged children, and 24.7 percent high-school aged children. I calculate the average effect (the dot product of the percent-in-group times the class size effect) which yields 0.012σ per ten-student change or 0.0012σ per student change.

Performance Pay

The evidence on performance pay suggests modest improvements to achievement in the presence of stronger incentives (Lavy 2002; Springer et al. 2010; Muralidharan and Sundararaman 2011; Sojourner, Fryer et al. 2011; Fryer 2013; Mykerezi, and West 2014; Dee and Wyckoff 2015; Imberman and Lovenheim 2015; Balch and Springer 2015). The settings of each study differ enough to make comparison difficult. In many programs, schools implemented the reform with other supports; in others, the incentives apply to school-wide or district-wide goals. Because of the program's similarity to the one posed to teachers in my survey and the setting is geographically proximate (from Houston, Texas), I use Imberman and Loveheim (2015) for a parameter value. They use the fact that grade-level incentives are stronger for smaller grades, and find that a $1,000 merit-pay increase induces a 0.0136σ increase in student achievement.

Highly rated teachers express stronger preferences for an offer containing merit pay than other teachers. To calculate the influence of performance pay on selection in retention, I simulate the retention patterns of a cohort of 10,000 hypothetical teachers and assume that they are uniformly distributed across ten quality deciles when they begin teaching (which conservatively assumes no positive selection into the teaching environment based on performance pay). I calculate the utility each of those teachers have for the compensation bundle for teaching, using the differential estimates of the top three deciles for performance pay, and I add a random component to their utility from the empirical distribution of the errors in the empirical model to reflect that estimated preferences are not deterministic. I then rank each teacher's utility for teaching from greatest to least so that I have an ordered set of teachers with the most prone to leave the profession at the bottom of the ranked set and the least likely arranged at the top.

Using the retention model constructed from Hendricks, I calculate what fraction of new teachers will attrit based on the considered compensation structure and working conditions. To construct the set of teachers who persists into a second year, I assume that those who attrit count up to that fraction of leavers from the bottom of the ranked set of teachers. (For example, if the Hendricks model predicts that 5 percent of new teachers will attrit, I copy the list of teachers from the first year to the second year while removing the 5 percent of teachers who had the lowest utility from teaching). Because the random component is substantial, those that least prefer teaching includes a substantial share of highly rated teachers, even when considering compensation bundles including significant in performance pay. I iterate this process for each year of a teacher's career to calculate, in the end, the distribution of types (what share of teachers are in each decile bin in steady state).

I allow the model to select whether to evaluate teachers using "VAM only" or "VAM and Danielson," a distinction that is important for calculating the impact of changing retention patterns. Using the teacher data, I calculate the average VA in each decile bin, controlling for teacher experience. That is, the performance pay program compares teachers to those with similar experience to reward talent, rather than experience, which is already rewarded by the salary gradient in experience on the salary schedule. (Interestingly, VA does not have a significant experience gradient in Aldine, but Danielson scores do). When creating deciles based on VAM and Danielson together, I normalize both VAMs and Danielson scores to have a SD equal to 1 and add the two measures together before generating decile bins based on the sum. I calculate the average VA in each decile bin based on VAM + Danielson and the average VA in each decile bin based on VAM alone, using only teacher observations that have both VAM and Danielson so the samples forming the VA vectors are identical. The dot product of the decile shares and these VA vectors generates the VA produced by the selection in retention of the considered compensation structure.

Online Appendix E: Online Appendix Figures

ONLINE APPENDIX FIGURE 1—SAMPLE COMPENSATION 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:	1.0% each year	2.0% each year
Reward:	Teachers receive $0 reward if they are in the top 25% of the school based on principal ratings and student growth	Teachers receive $0 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 with respect to compensation structure.

ONLINE APPENDIX FIGURE 2—SAMPLE WORKING-CONDITION 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:	$49,850	$52,700
Contract:	Teachers receive a renewable 3-year term contract after a 3-year probationary contract	Teachers receive a renewable 2-year term contract after a 1-year probationary contract
Distance from home:	15-minute drive	1-minute drive
Class size:	23	27
Assistance:	The school hires someone to help you with instructional support for 9 hours each week	The school hires someone to help you with instructional support for 0 hours each week
	○	○

Note: This figure presents an illustration of the questions answered by teacher respondents with respect to working conditions.

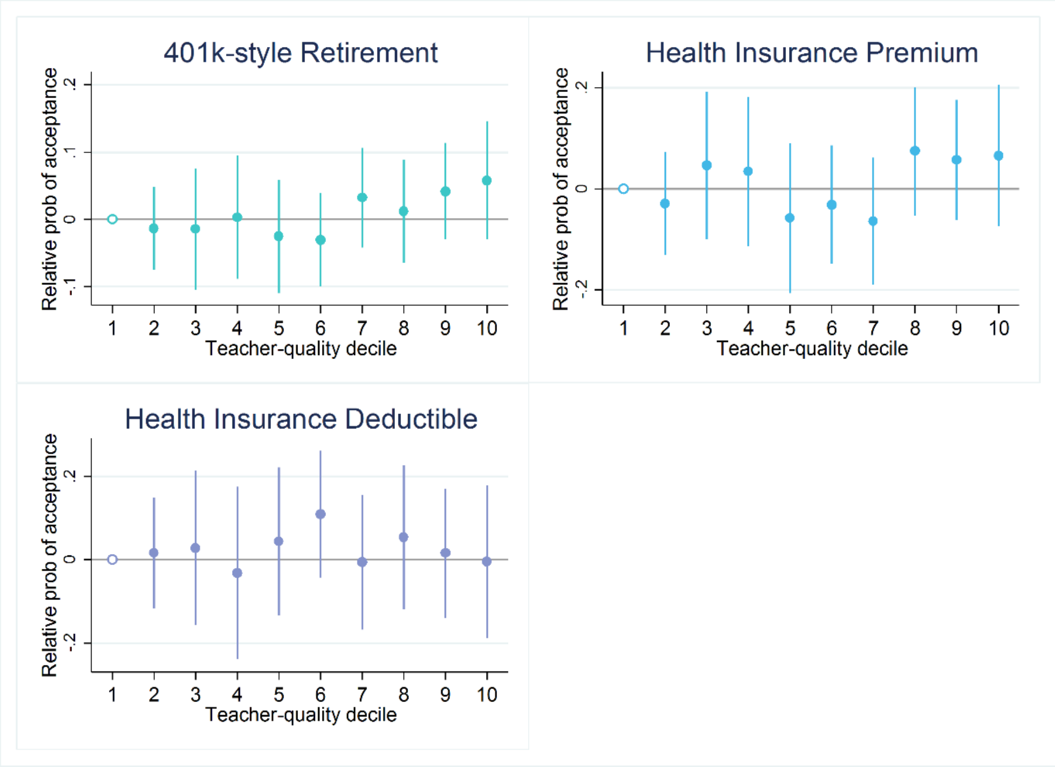
ONLINE APPENDIX FIGURE 3—SAMPLE STUDENTS-&-LEADERSHIP QUESTION

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

	School 1	**School 2**
Starting salary:	\$47,150	\$50,300
Percent of students in poverty:	38%	53%
Percent of students who are minority:	36%	66%
Average student achievement:	43rd percentile	57th percentile
Principal support:	Principals are hands-off with disruptive students	Principals are hands-off with disruptive students
School bus:	The school's buses are blue	The school's buses are not blue
	○	○

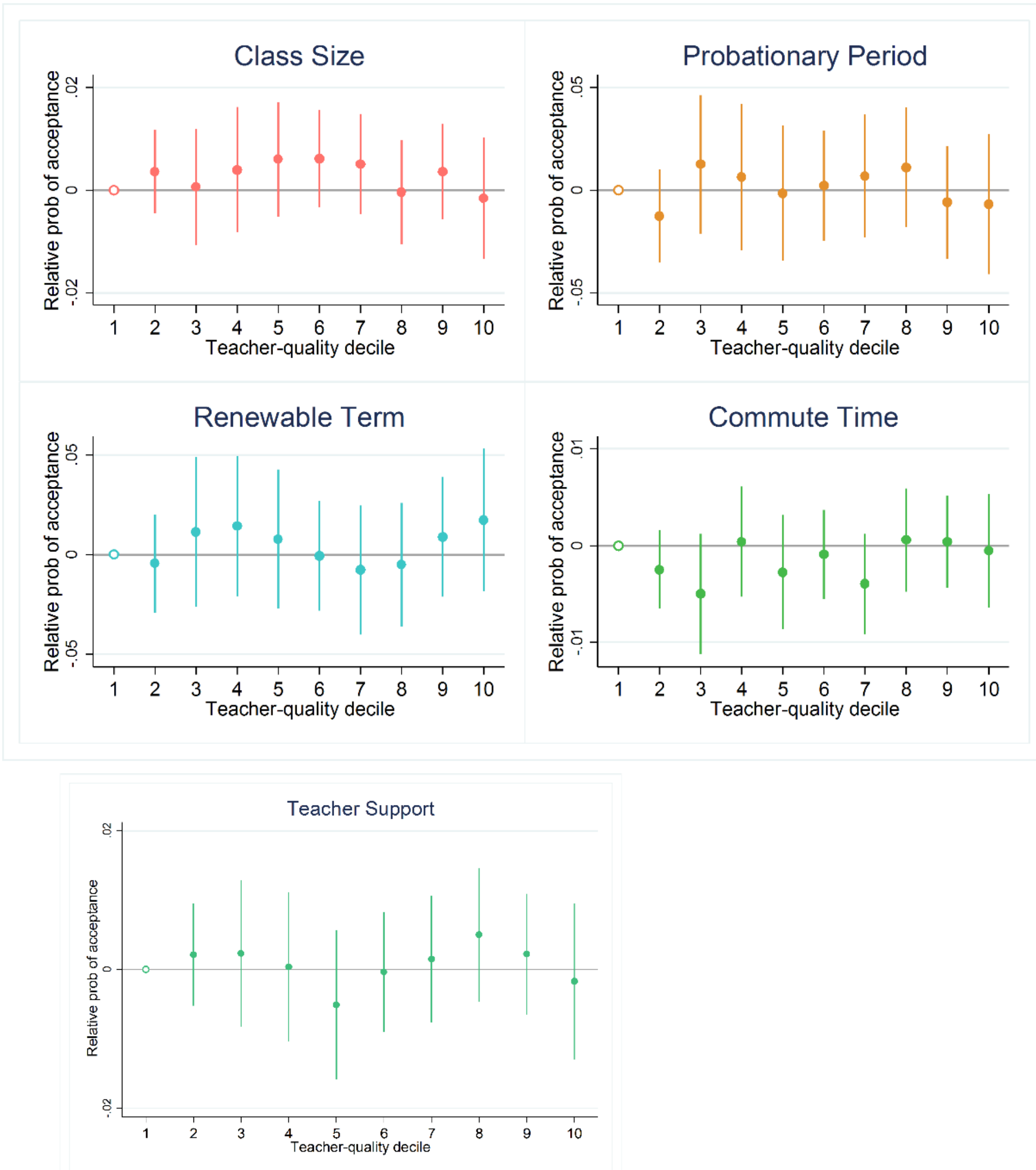
Note: This figure presents an illustration of the questions answered by teacher respondents with respect to student and principal characteristics.

ONLINE APPENDIX FIGURE 4—DIFFERENTIAL COMPENSATION PREFERENCE BY TEACHER-QUALITY DECILE



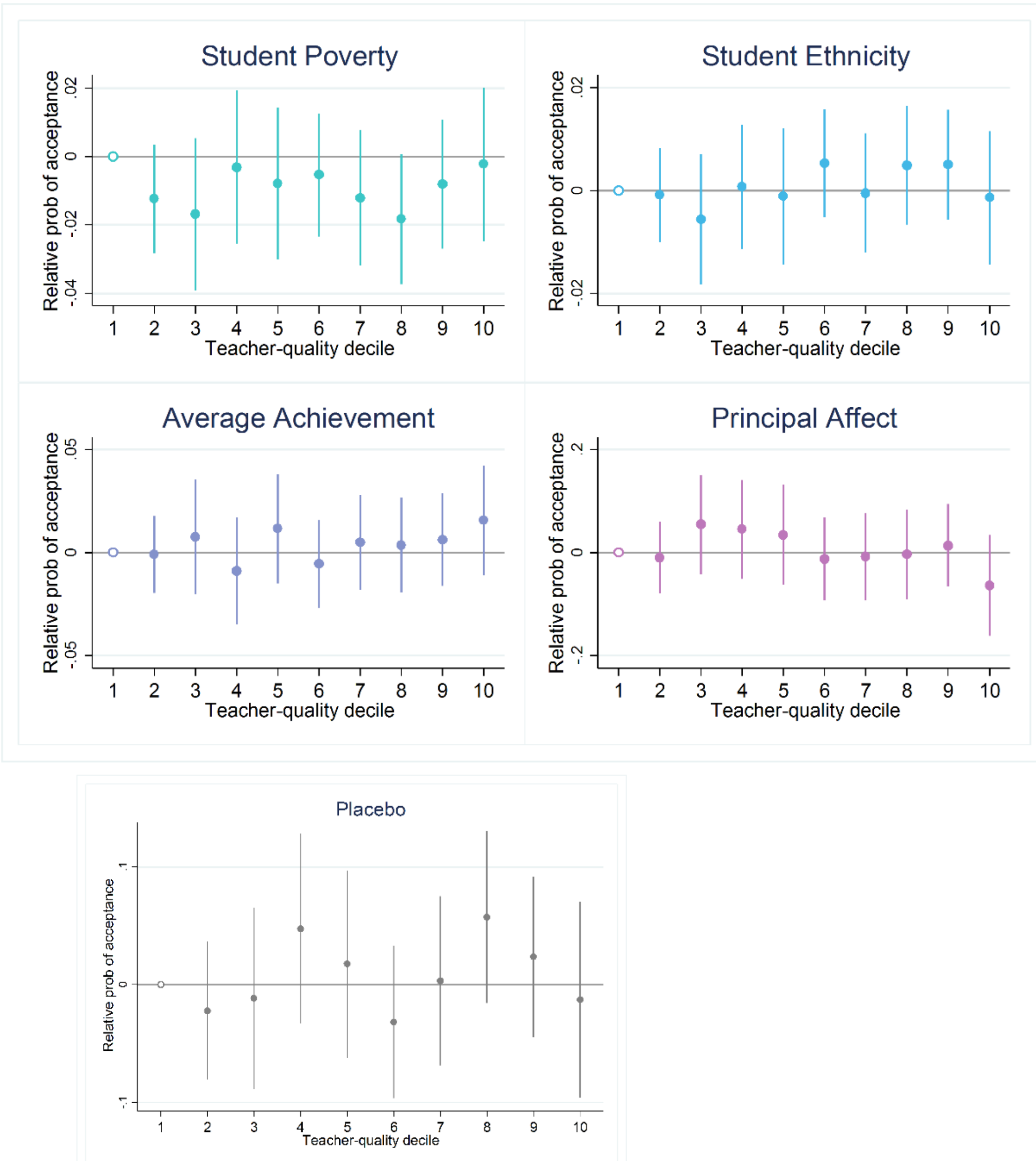
Note: This figure shows visually whether teachers of different quality deciles have distinct preferences for various compensation attributes, relative to bottom-decile teachers.

ONLINE APPENDIX FIGURE 5—DIFFERENTIAL WORKING-CONDITION PREFERENCE BY TEACHER-QUALITY DECILE



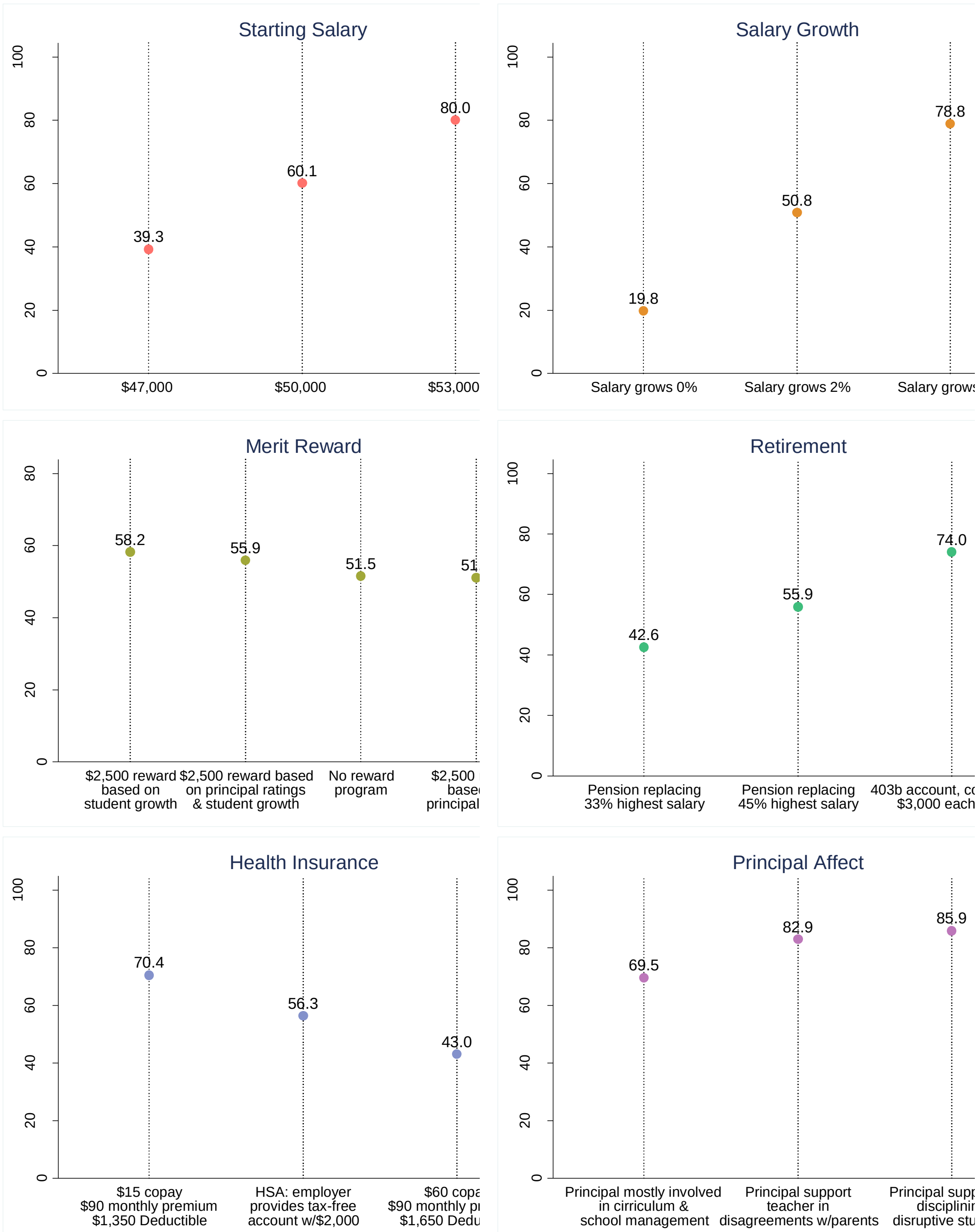
Note: This figure shows visually whether teachers of different quality deciles have distinct preferences for various working-condition attributes, relative to bottom-decile teachers.

ONLINE APPENDIX FIGURE 6—DIFFERENTIAL STUDENTS-&-LEADERSHIP PREFERENCE BY TEACHER-QUALITY DECILE



Note: This figure shows visually whether teachers of different quality deciles have distinct preferences for various student-and-leadership attributes, relative to bottom-decile teachers.

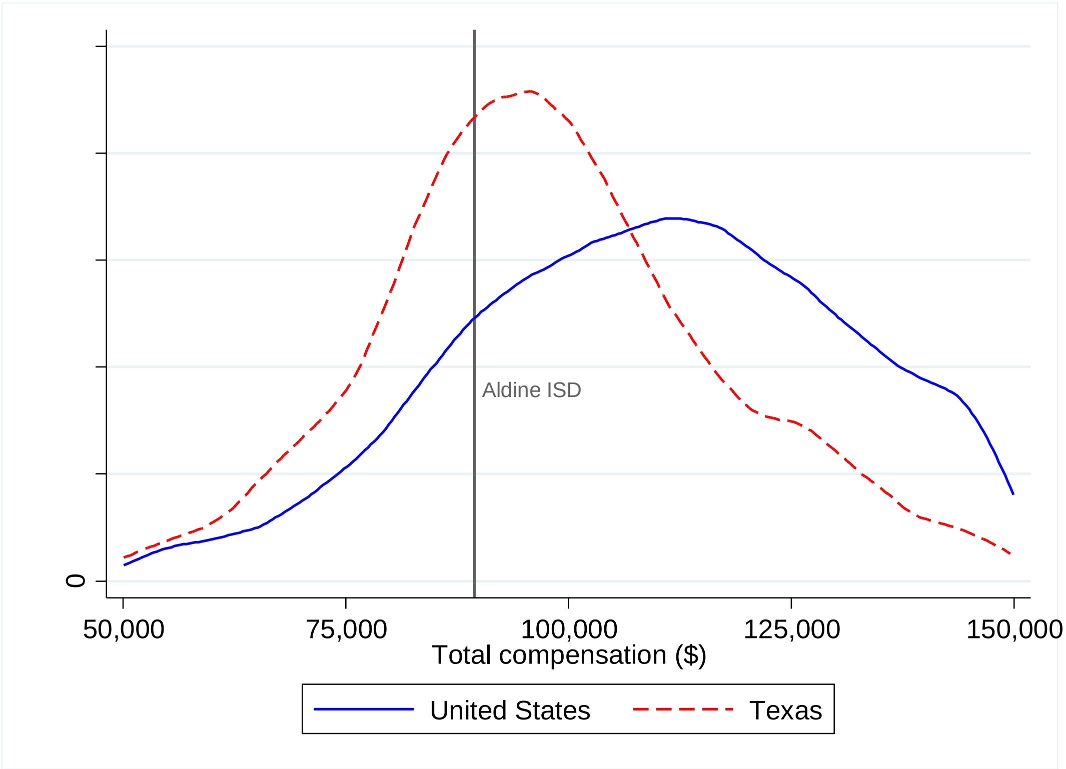
ONLINE APPENDIX FIGURE 7—STAND-ALONE ATTRIBUTE EVALUATION QUESTION



Note: This figure presents the results of additional survey questions in which a subset of teachers were asked to evaluate the probability that they would accept an offer that featured varying attributes.

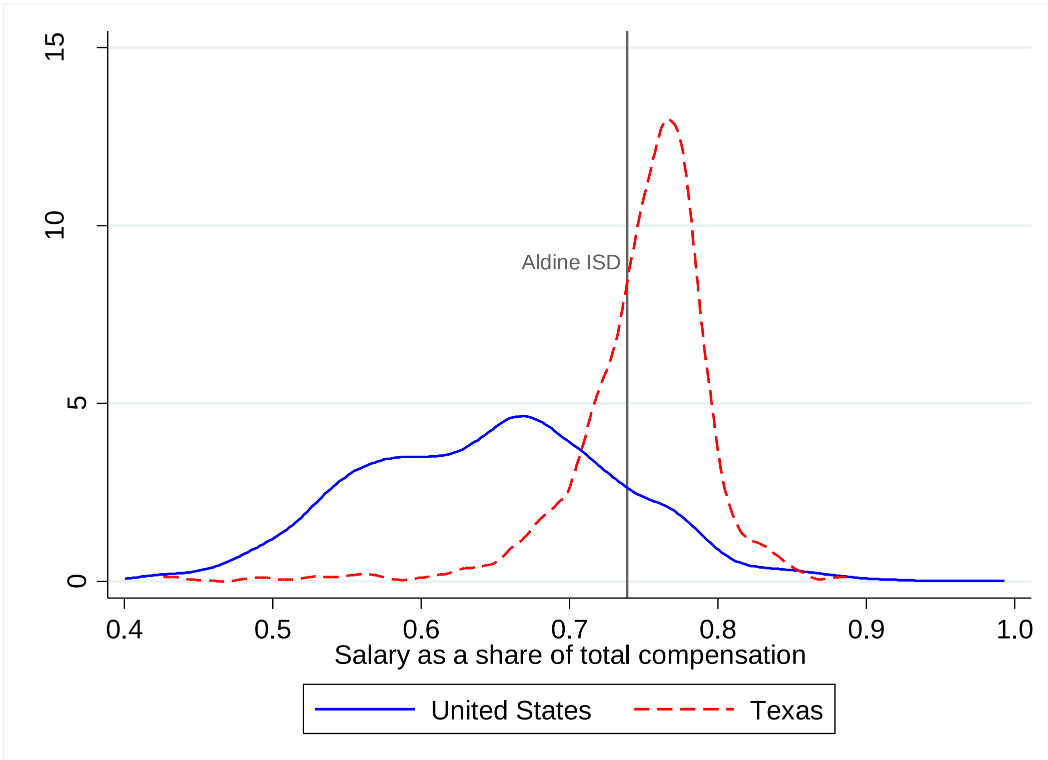
ONLINE APPENDIX FIGURE 8—COMPARING ALDINE-ISD

TOTAL COMPENSATION TO DISTRIBUTION



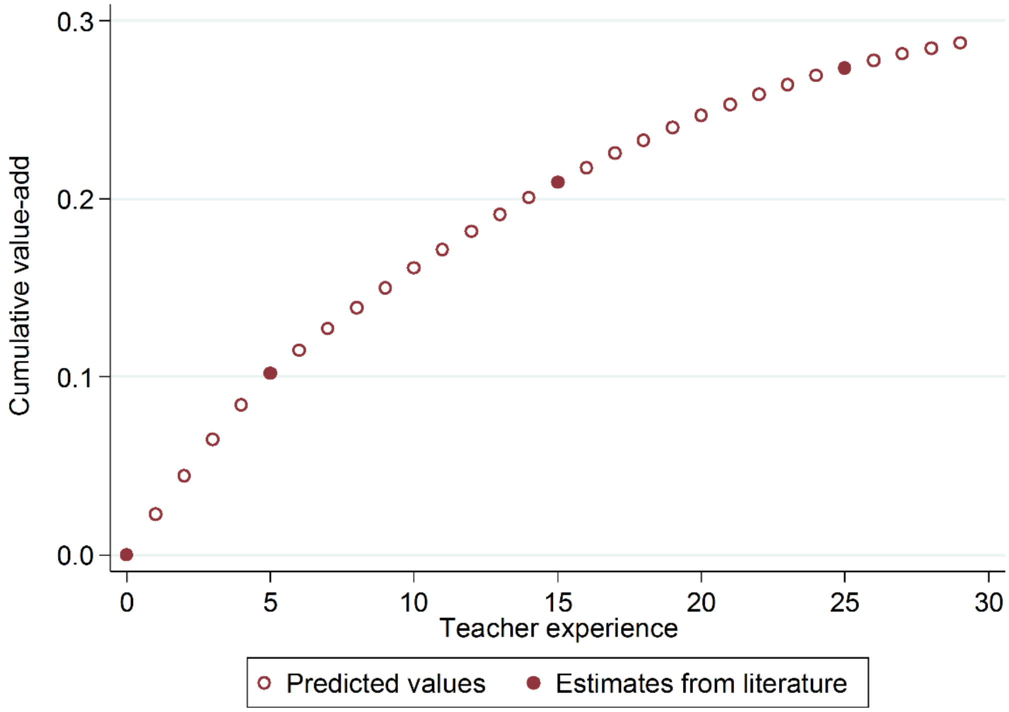
Note: This figure compares the average total compensation at Aldine ISD to the distribution of total compensation in the U.S. and in Texas using data from the Local Education Finance Survey.

ONLINE APPENDIX FIGURE 9—COMPARING ALDINE-ISD SALARY SHARE TO DISTRIBUTION



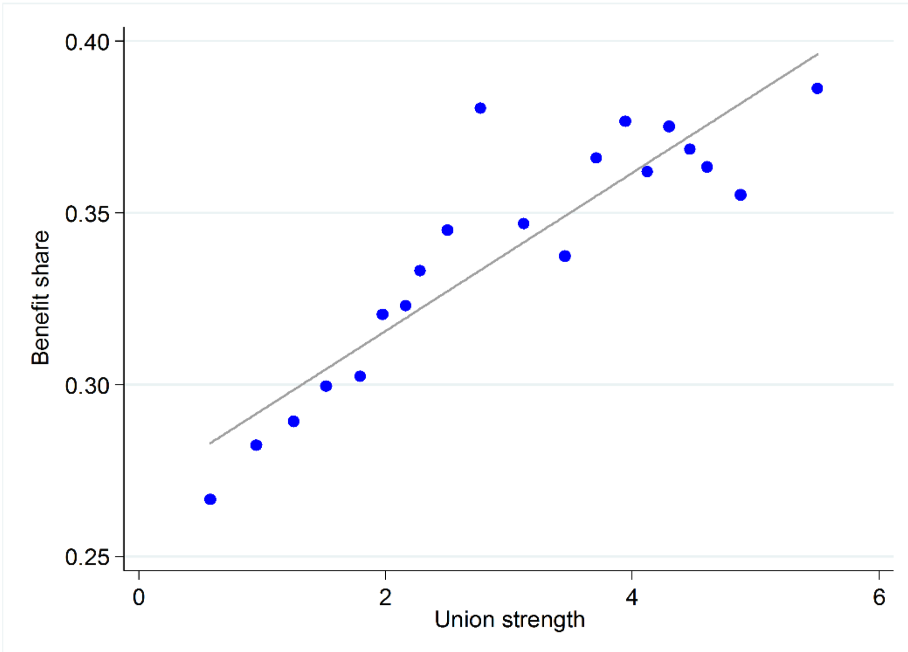
Note: This figure compares the average total compensation at Aldine ISD to the distribution of total compensation in the U.S. and in Texas using data from the Local Education Finance Survey.

ONLINE APPENDIX FIGURE 10—VALUE-ADDED GROWTH WITH EXPERIENCE



Note: This figure shows the value-added estimates from Papay and Kraft (2015) in the solid dots. The open dots represent the inferred value add for each experience level that I use in the achievement production function.

ONLINE APPENDIX FIGURE 11—UNION STRENGTH AND BENEFIT SHARE



Note: This figure shows the relationship between union strength and the share of a teacher's compensation that comes to her in the form of benefits, conditional on salary bins.

ONLINE APPENDIX F: ONLINE APPENDIX TABLES

ONLINE APPENDIX TABLE 1—OFFER ATTRIBUTES FOR CONJOINT EXPERIMENTS

Attribute	**Levels**
Salary	$46,550, $46,700, $46,850, $47,000, $47,150, $47,300…$53,300, $53,450
Growth	0.2%, 0.4%, 0.6%, 0.8%, 1.0%, 1.2%, 1.4%, 1.6%, 1.8%, 2.0%, 2.2%, 2.4%, 2.6%
Deductible	$1,200, $1,250, $1,300, $1,350, $1,400, $1,450, $1,500, $1,550, $1,600...$1,800
Premium	Monthly health insurance premium: $40, $90
Co-pay	$0, $5, $10, $15, $20, $25, $45, $50, $55, $60, $65, $70, $75
Reward	$0, $1,750, $2,000, $2,250, $2,500, $2,750, $3,000, $3,250
Rating	Evaluated based on: student growth and principal evaluations, student growth only
Retirement plan	pension, 403(b) (defined contributions)
Replacement rate	33%, 35%, 37%, 39%, 41%, 43%, 45%, 48%, 50%, 52%, 54%, ...63%, 65%, 67%
Time till tenure	immediate, 1 year, 2 years, 3 years
Review term	1 year, 2 years, 3 years, 4 years, 5 years
Commute time	1 minutes, 3 minutes, 5 minutes, 7 minutes, 9 minutes, 11 minutes...19 minutes
Hired assistance	0 hours per week, 5 hours per week, 7 hours per week, 9 hours per week
Poverty rate	38%, 43%, 47%, 48%, 53%, 58%, 63%, 68%, 72%, 77%, 82%...97%, 99%
Minority share	12%, 18%, 24%, 30%, 36%, 42%, 48%, 66%, 72%, 78%, 90%, 96%, 100%
Av. achmt prctle	percentiles: 23rd, 27th, 31st, 35th, 39th, 43rd, 47th, 53rd, 57th, 61st...73rd, 77th
Principal	hands-off with disruptive students, supportive with disruptive students
Bus color	blue, not blue

Note: This table presents all the possible values presented to respondents in the estimating sample.

ONLINE APPENDIX TABLE 2 – TEACHER DEMOGRAPHICS

	Average	Std. Dev.
Experience in years	9.03	(9.21)
Bachelor’s	0.455	(0.498)
Master’s	0.299	(0.458)
White	0.276	(0.447)
Hispanic	0.208	(0.406)
Black	0.367	(0.482)
Female	0.680	(0.467)
VAM score	0.000	(0.995)
Danielson score	12.8	(2.07)

Note: This table presents the demographic makeup of teacher respondents.

ONLINE APPENDIX TABLE 3— EFFECTS OF COMPENSATION ATTRIBUTES ON THE PROBABILITY THAT TEACHERS ACCEPT THE JOB OFFER (COMPLEMENT TO FIGURE 1)

	Choice		
	Coeff.	Std. err.	P-value
	(1)	(2)	(3)
Starting salary			
$51,000	0.266**	(0.010)	0.000
$54,000	0.460**	(0.015)	0.000
Salary growth			
1 percent	0.175**	(0.015)	0.000
2 percent	0.324**	(0.016)	0.000
Merit pay			
$2000	0.107**	(0.013)	0.000
$3000	0.062**	(0.012)	0.000
VAM only	-0.077**	(0.015)	0.000
Retirement			
Replaces 40%	0.095**	(0.022)	0.000
Replaces 50%	0.177**	(0.031)	0.000
Replaces 60%	0.381**	(0.022)	0.000
Replaces 70%	0.497**	(0.028)	0.000
401k-style	0.144**	(0.017)	0.000
Health insurance			
$50/mo. premium	0.048**	(0.009)	0.000
$1,300 deductible	0.018	(0.030)	0.544
R-squared	0.1904		
Adj. R-squared	0.1894		
Num. obs.	31,820		

Note: This table presents the estimates behind Figure 1. These results make bins to describe each of the attributes of available offers to show the influence of each characteristic nonparametrically.

ONLINE APPENDIX TABLE 4— EFFECTS OF WORKING-CONDITION ATTRIBUTES ON THE PROBABILITY THAT TEACHERS ACCEPT THE JOB OFFER (COMPLEMENT TO FIGURE II)

	Choice		
	Coeff.	Std. err.	P-value
	(1)	(2)	(3)
Class size			
24 students	-0.163**	(0.018)	0.000
28 students	-0.408**	(0.014)	0.000
Probationary period			
1-year	-0.084**	(0.021)	0.000
2-year	-0.072**	(0.019)	0.000
3-year	-0.190**	(0.021)	0.000
Renewable terms			
2-year	0.025*	(0.012)	0.047
3-year	-0.005	(0.010)	0.603
Commute time			
~10 minutes	-0.036**	(0.011)	0.001
~20 minutes	-0.075**	(0.011)	0.000
Teacher support			
5 hours/wk	0.169**	(0.011)	0.000
7 hours/wk	0.157**	(0.010)	0.000
9 hours/wk	0.188**	(0.011)	0.000
R-squared	0.281		
Adj. R-squared	0.280		
Num. obs.	31,574		

Note: This table presents the estimates behind Figure 2. These results make bins to describe each of the attributes of available offers to show the influence of each characteristic nonparametrically.

ONLINE APPENDIX TABLE 5— EFFECTS OF WORKING-CONDITION ATTRIBUTES ON THE PROBABILITY THAT TEACHERS ACCEPT THE JOB OFFER (COMPLEMENT TO FIGURE III)

	Choice		
	Coeff.	Std. err.	P-value
	(1)	(2)	(3)
Student poverty			
60% low-income	-0.017**	(0.019)	0.379
80% low-income	-0.081**	(0.017)	0.000
100% low-income	-0.116**	(0.023)	0.000
Student ethnicity			
60% minority	0.031	(0.019)	0.110
90% minority	0.012	(0.015)	0.429
Average achievement			
50th percentile	0.153**	(0.012)	0.000
66th percentile	0.253**	(0.030)	0.000
Principal affect			
Supportive	0.764**	(0.012)	0.000
Placebo			
Bus blue	0.009	(0.011)	0.402
R-squared	0.365		
Adj. R-squared	0.364		
Num. obs.	23,678		

Note: This table presents the estimates behind Figure 3. These results make bins to describe each of the attributes of available offers to show the influence of each characteristic nonparametrically.

ONLINE APPENDIX TABLE 6—PREFERENCES FOR WORKING CONDITIONS BY TEACHER QUALITY

	Choice		Choice	
	Reference Group	Quality-decile interaction	Reference Group	Quality-decile interaction
	(1)	(2)	(1)	(2)
Benchmark				
Starting salary	0.119**	-0.002	0.119**	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Contract				
Probationary period	-0.063**	0.011	-0.059**	0.010
	(0.008)	(0.012)	(0.008)	(0.012)
Term length	-0.009	0.015	-0.008	0.012
	(0.009)	(0.013)	(0.009)	(0.013)
Working conditions				
Commute time	-0.007**	0.002	-0.008**	0.002
	(0.001)	(0.002)	(0.001)	(0.002)
Class size	-0.071**	0.002	-0.072**	0.002
	(0.003)	(0.004)	(0.003)	(0.004)
Assistance	0.027**	0.001	0.028**	0.001
	(0.002)	(0.004)	(0.002)	(0.004)
Experience bins	X		X	
Exp. interactions	.		X	
R-squared	0.288		0.289	
Observations	21,312		21,312	

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 controls with experience bins interacted with each attribute.

ONLINE APPENDIX TABLE 7—PREFERENCES FOR STUDENT AND LEADERSHIP CHARACTERISTICS BY TEACHER QUALITY

	Choice		Choice	
	Reference Group (1)	Quality-decile interaction (2)	Reference Group (1)	Quality-decile interaction (2)
Benchmark				
Starting salary	0.068**	-0.002	0.068**	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Students				
Percent low income	-0.025**	0.002	-0.025**	0.002
	(0.005)	(0.008)	(0.005)	(0.008)
Percent minority	0.001	0.006	0.001	0.006
	(0.003)	(0.005)	(0.003)	(0.005)
Ave. achievement	0.027**	0.010	0.027**	0.010
	(0.005)	(0.009)	(0.005)	(0.009)
Principal affect				
Supportive	0.588**	-0.007	0.555**	-0.026
	(0.020)	(0.034)	(0.024)	(0.034)
Placebo				
Blue bus	-0.014	0.037	-0.026	0.034
	(0.017)	(0.028)	(0.020)	(0.029)
Experience bins	X		X	
Exp. interactions	.		X	
R-squared	0.373		0.375	
Observations	15,982		15,982	

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 controls with experience bins interacted with each attribute.

ONLINE APPENDIX TABLE 8—ASSESSING THE INFLUENCE OF DIFFERENT QUALITY MEASURES ON DIFFERENTIAL PREFERENCES FOR PERFORMANCE PAY

	Choice	Choice	Choice	Choice	Choice
	(1)	(2)	(3)	(4)	(5)
Reward	0.029**	0.023*	0.018**	0.019	0.013*
	(0.003)	(0.009)	(0.007)	(0.013)	(0.007)
Reward × VAM index		0.037**		0.036*	
		(0.014)		(0.018)	
Reward × Danielson index			0.032**	0.011	
			(0.012)	(0.018)	
Reward × Quality index					0.043**
					(0.010)
Observations	31,820	12,274	17,166	7,942	21,498

*Note: * $p < 0.05$, *** $p < 0.001$.* This table presents the interaction of merit pay with various teacher-quality indices; the results are qualitatively similar across the measure of quality we use.

ONLINE APPENDIX TABLE 9—EXPERIENCE HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability			
	Novice teachers (1st quartile: 0-1 yrs)	New-teacher differential (2nd quartile: 2-6 yrs)	Experienced-teacher differential (3rd quartile: 7-14 yrs)	Veteran-teacher differential (4th quartile: 15-36 yrs)
	(1)	(2)	(3)	(4)
Starting salary	0.093**	0.001	-0.009*	-0.029**
	(0.003)	(0.004)	(0.004)	(0.004)
Salary growth	0.205**	-0.019	-0.025*	-0.02
	(0.011)	(0.012)	(0.011)	(0.012)
Bonus amount	0.026**	0.009	0.013	-0.005
	(0.005)	(0.008)	(0.007)	(0.008)
VAM only	-0.077**	0.014	0.003	-0.012
	(0.017)	(0.018)	(0.018)	(0.017)
Replacement	0.012**	0.001	0.003*	0.006**
	(0.001)	(0.001)	(0.001)	(0.001)
401k-style	0.079**	-0.012	0.011	-0.014
	(0.014)	(0.020)	(0.020)	(0.020)
Premium (yearly)	-0.064*	-0.01	-0.013	-0.057
	(0.022)	(0.037)	(0.036)	(0.036)
Deductible	-0.589*	-0.062	0.265	0.965**
	(0.221)	(0.156)	(0.149)	(0.151)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying different levels of teacher experience. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 10—EXPERIENCE HETEROGENEITY IN WORKING-CONDITION PREFERENCES

	Linear Probability			
	Novice teachers	New-teacher differential	Experienced-teacher differential	Veteran-teacher differential
	(1st quartile: 0-1 yrs)	(2nd quartile: 2-6 yrs)	(3rd quartile: 7-14 yrs)	(4th quartile: 15-36 yrs)
	(1)	(2)	(3)	(4)
Probationary period	-0.045**	-0.007	0.003	0.002
	(0.005)	(0.007)	(0.006)	(0.006)
Term length	-0.003	-0.010	0.003	0.005
	(0.005)	(0.007)	(0.007)	(0.007)
Commute time	-0.005**	0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Class size	-0.054**	0.000	0.000	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)
Assistance	0.021**	0.000	0.004*	0.005*
	(0.001)	(0.002)	(0.002)	(0.002)

Note: $* p < 0.05$, $** p < 0.001$. This table presents a heterogeneity analysis by interacting each attribute with dummies identifying different levels of teacher experience. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 11—EXPERIENCE HETEROGENEITY IN STUDENT/PRINCIPAL PREFERENCES

	Linear Probability			
	Novice teachers	New-teacher differential	Experienced-teacher differential	Veteran-teacher differential
	(1st quartile: 0-1 yrs)	(2nd quartile: 2-6 yrs)	(3rd quartile: 7-14 yrs)	(4th quartile: 15-36 yrs)
	(1)	(2)	(3)	(4)
Percent low income	-0.031**	0.001	-0.001	0.010
	(0.005)	(0.007)	(0.007)	(0.007)
Percent minority	-0.001	0.003	0.006	0.009*
	(0.003)	(0.004)	(0.004)	(0.004)
Ave. achievement	0.048**	-0.006	-0.010	0.018*
	(0.005)	(0.009)	(0.008)	(0.008)
Supportive principal	0.722**	0.014	0.049	0.126**
	(0.020)	(0.032)	(0.030)	(0.029)
Blue bus	0.001	0.022	0.009	0.007
	(0.016)	(0.026)	(0.024)	(0.024)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying different levels of teacher experience. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 12—SEX HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability	
	Female teachers	Male differential
	(1)	(2)
Starting salary	0.082**	0.011**
	(0.002)	(0.003)
Salary growth	0.192**	-0.001
	(0.009)	(0.011)
Bonus amount	0.030**	-0.005
	(0.004)	(0.007)
VAM only	-0.079**	0.011
	(0.015)	(0.016)
Replacement	0.015**	0.000
	(0.001)	(0.001)
401k-style	0.084**	-0.035
	(0.011)	(0.018)
Premium (yearly)	-0.093**	0.053
	(0.016)	(0.033)
Deductible	-0.211	-0.513**
	(0.214)	(0.134)

Note: * $p < 0.05$, ** $p < 0.001$. This table presents a heterogeneity analysis by interacting each attribute with dummies identifying male teachers. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 13—SEX HETEROGENEITY IN WORKING-CONDITION PREFERENCES

	Linear Probability	
	Female teachers	Male differential
	(1)	(2)
Probationary period	-0.043**	-0.008
	(0.004)	(0.006)
Term length	-0.003	0.002
	(0.004)	(0.006)
Commute time	-0.005**	0.000
	(0.001)	(0.001)
Class size	-0.055**	0.007**
	(0.001)	(0.002)
Assistance	0.025**	-0.008**
	(0.001)	(0.002)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying male teachers. Standard errors clustered at the teacher level.

Online Appendix Table 14—Sex Heterogeneity in Student and Principal Preferences

	Linear Probability	
	Female teachers	Male differential
	(1)	(2)
Percent low income	-0.027**	-0.005
	(0.003)	(0.006)
Percent minority	0.004*	-0.001
	(0.002)	(0.004)
Ave. achievement	0.048**	0.000
	(0.004)	(0.008)
Supportive principal	0.792**	-0.130**
	(0.013)	(0.027)
Blue bus	0.015	-0.028
	(0.012)	(0.022)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying male teachers. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 15—RACIAL HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability		
	White teachers	Black differential	Hispanic differential
	(1)	(2)	(3)
Starting salary	0.082**	0.004	0.008*
	(0.003)	(0.003)	(0.004)
Salary growth	0.213**	-0.048**	-0.017
	(0.011)	(0.010)	(0.011)
Bonus amount	0.011*	0.037**	0.023*
	(0.005)	(0.006)	(0.007)
VAM only	-0.086**	0.028	0.005
	(0.016)	(0.015)	(0.017)
Replacement	0.016**	-0.001	-0.002
	(0.001)	(0.001)	(0.001)
401k-style	0.059**	0.035*	0.024
	(0.013)	(0.016)	(0.019)
Premium (yearly)	-0.077**	-0.002	-0.02
	(0.021)	(0.030)	(0.035)
Deductible	-0.239	-0.067	-0.247
	(0.221)	(0.127)	(0.148)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying black teachers and Hispanic teachers. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 16—RACIAL HETEROGENEITY IN WORKING-CONDITION PREFERENCES

	Linear Probability		
	White teachers	Black differential	Hispanic differential
	(1)	(2)	(3)
Probationary period	-0.037**	-0.021**	-0.003
	(0.005)	(0.005)	(0.006)
Term length	0.002	-0.014*	0.000
	(0.005)	(0.006)	(0.007)
Commute time	-0.006**	0.001	0.001
	(0.001)	(0.001)	(0.001)
Class size	-0.055**	0.007**	-0.005*
	(0.001)	(0.002)	(0.002)
Assistance	0.023**	0.001	-0.001
	(0.001)	(0.002)	(0.002)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying black teachers and Hispanic teachers. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 17—RACIAL HETEROGENEITY IN STUDENT AND PRINCIPAL PREFERENCES

	Linear Probability		
	White teachers	Black differential	Hispanic differential
	(1)	(2)	(3)
Percent low income	-0.031**	0.008	-0.002
	(0.004)	(0.006)	(0.007)
Percent minority	0.000	0.011*	-0.001
	(0.003)	(0.003)	(0.004)
Ave. achievement	0.058**	-0.021*	-0.008
	(0.005)	(0.007)	(0.008)
Supportive principal	0.809**	-0.065*	-0.099**
	(0.017)	(0.024)	(0.030)
Blue bus	0.013	-0.014	0.005
	(0.014)	(0.020)	(0.024)

*Note: * p < 0.05, ** p < 0.001.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying black teachers and Hispanic teachers. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 18—LEAVER HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability		Linear Probability	
	Teachers that stay	Marginal-teacher differential	Teachers that stay	Marginal-teacher differential
	(1)	(2)	(3)	(4)
Starting salary	0.085**	-0.002	0.087**	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Salary growth	0.186**	0.008	0.193**	0.011
	(0.010)	(0.010)	(0.010)	(0.010)
Bonus amount	0.031**	0.003	0.035**	0.004
	(0.004)	(0.006)	(0.004)	(0.006)
VAM only	-0.068**	-0.017	-0.069**	-0.009
	(0.016)	(0.015)	(0.019)	(0.016)
Replacement	0.014**	0.001	0.014**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
401k-style	0.085**	-0.023	0.097**	-0.015
	(0.012)	(0.016)	(0.016)	(0.017)
Premium (yearly)	-0.095**	0.027	-0.088**	0.025
	(0.027)	(0.027)	(0.017)	(0.027)
Deductible	-0.252	0.006	-0.124	0.002
	(0.225)	(0.037)	(0.228)	(0.037)
Experience bins	X		X	
Exp. interactions	.		X	

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying teachers who left shortly after the survey, while nonparametrically controlling for teaching experience in yearly bins. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 19—LEAVER HETEROGENEITY IN WORKING CONDITION PREFERENCES

	Linear Probability		Linear Probability	
	Teachers that stay	Marginal-teacher differential	Teachers that stay	Marginal-teacher differential
	(1)	(2)	(3)	(4)
Probationary period	-0.049**	0.009	-0.047**	0.008
	(0.004)	(0.005)	(0.005)	(0.005)
Term length	0.000	-0.006	-0.001	-0.007
	(0.005)	(0.006)	(0.005)	(0.006)
Commute time	-0.004**	-0.001	-0.005**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Class size	-0.055**	0.004*	-0.055**	0.004*
	(0.001)	(0.002)	(0.001)	(0.002)
Assistance	0.022**	0.003*	0.022**	0.003*
	(0.001)	(0.002)	(0.001)	(0.002)
Experience bins	X		X	
Exp. interactions	.		X	

*Note: * p < 0.05, ** p < 0.001.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying teachers who left shortly after the survey, while nonparametrically controlling for teaching experience in yearly bins. Standard errors clustered at the teacher level.

0

ONLINE APPENDIX TABLE 20—LEAVER HETEROGENEITY

IN STUDENT AND PRINCIPAL PREFERENCES

	Linear Probability		Linear Probability	
	Teachers that stay	Marginal-teacher differential	Teachers that stay	Marginal-teacher differential
	(1)	(2)	(1)	(2)
Percent low income	-0.028**	0.000	-0.029**	0.000
	(0.004)	(0.006)	(0.004)	(0.006)
Percent minority	0.004	0.002	0.004	0.002
	(0.002)	(0.003)	(0.002)	(0.003)
Ave. achievement	0.043**	0.011	0.043**	0.011
	(0.004)	(0.007)	(0.004)	(0.007)
Supportive principal	0.760**	0.014	0.709**	-0.018
	(0.015)	(0.024)	(0.024)	(0.026)
Blue bus	0.006	-0.007	-0.016	-0.010
	(0.013)	(0.021)	(0.019)	(0.022)
Experience bins	X		X	
Exp. interactions	.		X	

Note: * $p < 0.05$, ** $p < 0.001$. This table presents a heterogeneity analysis by interacting each attribute with dummies identifying teachers who left shortly after the survey, while nonparametrically controlling for teaching experience in yearly bins. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 21—GRADE-LEVEL HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability		
	Elementary School	Middle School	High School
	(1)	(2)	(3)
Starting salary	0.090**	0.002	0.001
	(0.003)	(0.004)	(0.004)
Salary growth	0.193**	0.003	-0.007
	(0.012)	(0.012)	(0.013)
Bonus amount	0.035**	-0.001	-0.017*
	(0.006)	(0.008)	(0.008)
VAM only	-0.074**	0.010	0.011
	(0.019)	(0.018)	(0.019)
Replacement	0.014**	0.000	0.000
	(0.001)	(0.001)	(0.001)
401k-style	0.079**	-0.010	0.011
	(0.015)	(0.021)	(0.022)
Premium (yearly)	-0.061*	0.009	-0.07
	(0.025)	(0.038)	0.039
Deductible	-0.286	-0.082	0.043
	(0.167)	(0.156)	(0.167)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying school type. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 22— GRADE-LEVEL HETEROGENEITY IN WORKING-CONDITION PREFERENCES

	Linear Probability		
	Elementary School	Middle School	High School
	(1)	(2)	(3)
Probationary period	-0.038**	-0.017*	-0.022*
	(0.005)	(0.006)	(0.007)
Term length	0.006	-0.010	-0.014
	(0.006)	(0.012)	(0.007)
Commute time	-0.004**	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Class size	-0.062**	0.011**	0.016**
	(0.002)	(0.002)	(0.002)
Assistance	0.023**	0.001	-0.004
	(0.001)	(0.002)	(0.002)

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents a heterogeneity analysis by interacting each attribute with dummies identifying school type. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 23— GRADE-LEVEL HETEROGENEITY IN STUDENT AND PRINCIPAL PREFERENCES

	Linear Probability		
	Elementary School	Middle School	High School
	(1)	(2)	(3)
Percent low income	-0.029**	-0.004	-0.008
	(0.005)	(0.007)	(0.007)
Percent minority	0.000	0.006	0.008
	(0.003)	(0.004)	(0.004)
Ave. achievement	0.038**	0.004	0.012
	(0.006)	(0.008)	(0.009)
Supportive principal	0.757**	0.034	0.012
	(0.022)	(0.031)	(0.033)
Blue bus	0.023	-0.011	-0.057*
	(0.018)	(0.025)	(0.027)

Note: * $p < 0.05$, ** $p < 0.001$. This table presents a heterogeneity analysis by interacting each attribute with dummies identifying school type. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 24—COMPENSATION ESTIMATES FOR SIMULATION EXERCISES

	Linear (1)	Quadratic (2)
Starting salary	0.0846**	0.2863*
	(0.0022)	(0.1376)
Starting sal. sqr.		-0.0020
		(0.0014)
Salary grth.	0.1918**	0.2225**
	(0.0091)	(0.0370)
Salary grth. sqr.		-0.0145
		(0.0136)
Performance pay	0.0293**	0.1326**
	(0.0034)	(0.0232)
Performance pay sqr.		-0.0386**
		(0.0085)
VAM only	-0.0767**	-0.0699**
	(0.0145)	(0.0175)
Retirement replcmnt.	0.0146**	0.0388**
	(0.0005)	(0.0077)
Retire. replmt. sqr.		-0.0002*
		(0.0001)
401k-style	0.0767**	0.0524**
	(0.0100)	(0.0135)
Deductible	-0.3117	-0.3003
	(0.2115)	(0.2335)
Premium	-0.0821**	-0.1000**

	(0.0141)	(0.0160)
Observations	0.193	0.195
R-squared	31,820	31,820

*Note: * p < 0.05, ** p < 0.001*. This table presents the estimated utility coefficients for the simulation exercises; standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 25—WORKING CONDITIONS ESTIMATES FOR SIMULATION EXERCISES

	Linear	Quadratic
	(1)	(2)
Starting salary	0.0846**	0.0787**
	(0.0013)	(0.0016)
Time-to-tenure	-0.0424**	-0.0450**
	(0.0036)	(0.0037)
Review frequency	-0.0028	-0.0065
	(0.0037)	(0.0037)
Commute time (mins)	-0.0045**	-0.0026**
	(0.0005)	(0.0006)
Class size	-0.0502**	0.0916*
	(0.0011)	(0.0289)
Class size sqr.		-0.0029**
		(0.0006)
Assistance (hrs/wk)	0.0217**	0.0351**
	(0.0008)	(0.0039)
Assistance sqr.		-0.0018**
		(0.0005)
Observations	0.279	0.281
R-squared	31,574	31,574

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents the estimated utility coefficients for the simulation exercises; estimates are adjusted so that they are directly comparable to the coefficient estimates in prior table. Standard errors clustered at the teacher level.

ONLINE APPENDIX TABLE 26—RELATIONSHIP BETWEEN UNION INFLUENCE AND BENEFIT SHARE

	Benefit share (1)	Benefit share (2)	Benefit share (3)
Union strength	0.0260**	0.0274**	0.0278**
	(0.006)	(0.007)	(0.007)
Salary level	No	Yes	No
Salary bins	No	No	Yes
Mean DV	0.355	0.355	0.355
Observations	14,389	14,389	14,389
R-squared	0.187	0.192	0.268

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents the relationship between union strength and the share of a teacher's compensation received in benefits. Data from LEFS; standard errors clustered at the state level.

Online Appendix G: Invitation Materials

EXHIBIT 1—EMAIL INVITATION

Subject: Annual Aldine Survey for [David] (Amazon Gift Card as Payment)

Dear [David],

Please take a few minutes to respond to Aldine's annual survey. Your insights will be really helpful as we improve [MacArthur Elementary] policies to meet your needs.

Follow this link to the Survey: Take the Survey

Or copy and paste the URL below into your internet browser:
https://wharton.qualtrics.com/SE?Q_DL=2hF0dupiHRHRNPL_3kOTLRFe6J82Uy9_MLRP_eajfSglx4nJPpOt&Q_CHL=email

Participating in this survey means answering a series of questions about your experience in an Aldine ISD school. It will take about 15 minutes to complete. All information will be kept strictly confidential and no Aldine ISD employee will have access to your individual responses. **If you take the survey in the next three days, you have a chance to win one of 75 $10 Amazon gift certificates as payment which we will email to you directly!**

Please feel free to contact me with any comments or questions.

Thank you so much for all your help and all you do!

Thanks so much!
Andrew
Researcher, University of Pennsylvania

Follow this link to the Survey:
Take the Survey
Or copy and paste the URL below into your internet browser:
https://wharton.qualtrics.com/SE?Q_DL=2hF0dupiHRHRNPL_3kOTLRFe6J82Uy9_MLRP_eajfSglx4nJPpOt&Q_CHL=email
Follow the link to opt out of future emails:

Click here to unsubscribe

EXHIBIT 2—OVERVIEW OF TEACHER SURVEY

Overview of Teacher Survey

Contact Information: University of Pennsylvania, 3700 Market Street, Philadelphia, PA 19104, johnsta@upenn.edu

What is the purpose of the study and what will you be asked to do?
The purpose of the study is to learn more about the attitudes and experiences of teachers. Your insights will contribute to the improvement of Aldine ISD's teacher policies and training. Participating in this study entails answering a series of questions about your attitudes and experiences toward your work. The survey will take no more than 10-15 minutes to complete.

How will confidentiality be maintained and your privacy be protected?

Your participation in this study is voluntary. The research team will make every effort to keep all the information you tell us during the study strictly confidential, as required by law. The Institutional Review Board (IRB) at the University of Pennsylvania is responsible for protecting the rights and welfare of research volunteers like you. We have assigned you a confidential ID number, and any information you provide will be stored using that ID number. Separately, we maintain a key linking ID to name, and this key is stored in a separate file on a password-protected server at the University of Pennsylvania. All data collected in the study will be kept strictly confidential and separate from official Aldine ISD records. No Aldine ISD staff member will have access to your individual responses.

What should you do if you have questions?

If you have questions about the survey or your participation in this study, please email johnsta@upenn.edu.

By completing the following web pages, you are agreeing to take part in the research study. Thank you very much for your participation.

EXHIBIT 3—PRE-QUESTION INSTRUCTIONS

The first 11 questions will ask you to choose between two hypothetical job offers. We thank you for carefully considering each offer and designating which offer you would prefer.