Teacher Utility, Separating Equilibria, and Optimal Compensation:

Evidence from a Discrete-Choice Experiment

30 October 2018

By Andrew C. Johnston[1]

Improving schools depends largely on attracting and retaining high-quality teachers, a means made straightforward by understanding teacher preferences for compensation structure and costly amenities. Since these preferences cannot be estimated from naturally occurring choice data, I deploy a discrete-choice experiment where teachers have reason to reveal their preferences. Estimated utility suggests that districts overpay in retirement and class size while underpaying in salary and merit rewards. If the best teachers have distinct preferences, compensation can be structured to differentially attract and retain them; high-quality teachers have preferences like other teachers for most attributes but have stronger preferences for schools offering merit rewards.

I. Introduction

Teachers are the central factor by which the state fosters human-capital formation (Rivkin, Hanushek, and Kain 2005; Mansfield 2015). Great teachers endow their students with higher achievement, non-cognitive skills, and better long-run outcomes including higher earnings, less teenage fertility, and greater health (Chetty, Friedman, and Rockoff 2014).[2] In this light, it's unfortunate that some researchers find that teacher quality has declined markedly over the past half century (Corcoran, Evans, and Schwab 2002; Bacolod 2007).

Reversing, or even stanching, this trend has proven difficult for several reasons. On the demand side, it is challenging for schools to identify the best prospective teachers when hiring (Rockoff et al. 2011; Jacob et al. 2018) and known training programs are largely ineffective at

[1] Johnston: Department of Economics, University of California, Merced; email: acjohnston@ucmerced.edu. I am grateful to Dale Ballou, Barbara Biasi, Damon Clark, Sean Corcoran, Julie Cullen, Gordon Dahl, Laura Giuliano, Thomas Kane, Cory Koedel, Alexandre Mas, Rebecca Maynard, Richard Murnane, Michael Podgursky, Jonah Rockoff, Kenneth Shores, Christopher Walters, and seminar participants at APPAM and the Wharton School for keen feedback. I would like to thank Benjamin Feis for outstanding research assistance. Financial support from the Institute for Education Science, the U.S. Department of Labor's Employment, 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.

[2] The impacts of better teachers appear to be wide-ranging. For instance, being exposed to a teacher with σ 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, are more likely to be married, and save more for retirement.

improving teacher quality (Rockoff 2004; Rivkin et al. 2005; Kane et al. 2008; Harris and Sass 2011). On the supply-side, the profession is taxing but provides only middling compensation (Ingersoll 2003); at the same time, rigid pay schedules may lead to negative selection (Stinebrickner 2001; Hoxby and Leigh 2004; Biasi 2017), especially if the best teachers have attractive options outside of the teaching profession (Murnane and Olsen 1989; Bacolod 2007; Feng 2010; Chingos and West 2012; Wiswall 2013). Comprehending the teacher labor-supply decision, structured by preferences, illuminates how best to attract and retain excellent teachers with limited resources.

Empirically, estimating teacher preferences presents a challenge. Equilibrium matching between workers and jobs reflect not only candidate preferences but also labor-market conditions and employer choice (Wiswall and Zafar 2017). Teacher preferences could be disentangled by constructing choice sets from which teachers made their employment decisions (Train 2009). Unfortunately, records needed to construct menus comprising each teacher's options (concurrent job offers) do not exist (compare, for example, Avery et al. 2013).[3] Even if these records were attainable, however, they would be uninformative for a variety of reasons. First, observed characteristics in realized data are likely correlated with unobserved characteristics, confounding results. Second, even if the records existed and omitted variables could be somehow addressed, the variation needed to separate preferences for various components of compensation, contract, and working conditions is extremely limited, and ultimately insufficient, since teacher contracts are largely homogeneous across employers with many important attributes being expressly colinear or everywhere absent.[4]

[3] I approached a number of districts and they do not keep records of job offers made. But the more promising approach was that I approached a number of software firms that provide HR programs to schools and school districts in large consortiums. Unfortunately, although the software provided for offers to be made through the software, fewer than 1% of schools did. Similarly, teachers could accept offers through the software, but an even smaller fraction did so.

[4] State policy and common union influence generate similar compensation structures across districts. Within district, compensation is totally uniform. Many states provide a uniform pension and health insurance program, rendering teacher choice uninformative as it relates to compensation structure. Importantly, real-world data are particularly unhelpful in determining preferences for merit pay or alternative retirement vehicles which vary rarely. 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.

To address these challenges, I employ a choice experiment that permits the estimation of teacher preferences for compensation structure, contract type, and working conditions. Teachers in a large, urban school district were presented a series of hypothetical job offers, among which they select their preferred setting. The description of each choice set is constructed to require teachers to make tradeoffs between valued features including salary or retirement generosity, larger merit rewards, smaller class sizes, and expedited time-to-tenure. Importantly, the survey was delivered through an organization hired to update the district's compensation structure, providing a credible reason for teachers to thoughtfully consider and truthfully reveal their preferences. There are several other features that recommend the setting: the response rate was high (97.8 percent), there is evidence that inattention is not a significant concern, and measurement error in respondent choice (i.e., mistakes) will not lead to bias in the parameter estimates so long as mistakes are independent of the attributes (Wooldridge 2010). The resulting choice data allow for preference estimation over several facets of the work setting, which has not been feasible to date.

Teachers appear consistent and sophisticated. For instance, their utility is monotonic in salary and retirement, and they value health-insurance subsidies almost identically to an equivalent increase in salary, which is remarkable because these features are presented in different units (monthly premia versus yearly salaries). Moreover, teachers regard merit rewards approximately equal to their expected payout. The willingness-to-pay (WTP) estimates suggest teachers value a ten-student class-size reduction equal to a $4,880 increase in salary (9.8 percent of base pay),[5] seven times less than the cost such a reduction. In this setting, teachers consistently prefer plausibly 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 $415

[5] Here, base starting pay is $50,000 for a new teacher without a master's degree.

(0.8 percent of base pay), and teachers prefer schools with fewer students in poverty and higher academic achievement.

I explore whether the best teachers have distinct preferences, which could prove useful to policy makers. As stated, it is difficult to discern *ex ante* which prospective teachers will be highly productive (Hanushek 1986, 1997; Greenwald et al. 1996; Rockoff et al. 2011). If high-type teachers have distinct preferences for work aspects controlled by policy, however, policymakers can construct a separating equilibrium by structuring compensation, contracts, or working conditions to conform to the preferences of high performers. Under such a policy, high-type teachers might naturally select into the teaching setting, and low-type teachers are less likely to apply, a complement, or substitute, to costly screening.[6] Over time, moreover, high-type teachers are more likely to be retained, while the attrition of low-type teachers is likely unaffected, allowing additional draws for quality teachers (Ballou 1996; Hanushek 2011).[7] Using value-added models and principal evaluations, I find that the best teachers have broadly similar preferences to their peers, except in one regard: excellent teachers systematically prefer offers that include the opportunity to earn merit rewards. For instance, high-quality teachers (top decile) are 21 percent more likely than a low-quality teacher (bottom decile) to select an offer providing $3,000 in merit pay.[8] This analysis is necessarily partial equilibrium, and it's unclear, at least from this setting, the implications of teacher sorting into the teaching profession.

The preference estimates allow us to explore the optimal structure of compensation and working conditions under various objective functions. I estimate teacher utility functions with diminishing marginal returns and calibrate cost and achievement functions. When maximizing teacher utility, teacher retention, or the achievement production function subject to the existing

[6] Over time the effect may be especially pronounced since the preferred compensation differentially attracts and retains high-functioning teachers who also prefer work settings inhabited by other high-caliber colleagues (Feng and Sass 2016).
[7] In particular, 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.
[8] This finding holds when using the measure that teachers in the district are never informed of, their value-add measure, which suggests that highly able teachers have a sense of their ability even before measurement.

budget constraint, I find that, in each case, teachers are overpaid in benefits and working conditions and underpaid in salary and merit rewards. Simply restructuring what teachers are currently paid to reflect their preferences would yield an effective raise of $3,432, generating a 2.2 percent increase in teacher welfare. Restructuring pay to maximize student achievement would, in the long run, yield increases of 0.23σ, while having a slight, positive effect on teacher welfare. About half of the increase in achievement stems from improved selection due to merit rewards and a quarter emanates from improved retention arising from more spending on salary for experienced teachers.

Many aspects of this paper are distinct, but the study grows from the literatures on teacher preferences (Antos and Rosen 1975; Ballou 1996; Boyd et al. 2013; Biasi 2017), teacher compensation (Hanushek 1986; Card and Krueger 1992; Ballou and Podgursky 1997; Figlio 1997; Loeb and Page 2000; Hendricks 2014), and teacher quality (Rockoff 2004; Hanushek and Rivkin 2006; Chetty, Rockoff, and Friedman 2014). Previous studies have largely relied on equilibrium data to estimate preferences, inheriting a host of confounding factors. Due to data limitations, moreover, prior studies are not able to estimate the perceived value of various components of their compensation structure and working conditions, usually only salary. This study advances prior work by creating a transparent choice environment in which I measure teacher preferences over several important elements of the work setting, including dimensions for which there would be insufficient variation in naturally occurring records. It is the first paper to use choice data to calculate optimal compensation structure, I believe, in any profession. Finally, this paper advances the discussion of whether merit pay may be an effective tool for policy makers, not only by inducing greater effort[9] but also by influencing teacher selection (Lazear 2000, 2003; Rothstein 2015). The project represents an

[9] See Lavy 2002; Fryer 2011; Muralidharan and Sundararaman 2011; Dee and Wyckoff 2015; Imberman and Lovenheim 2015

effort to optimally structure compensation and working conditions to attract and retain talented teachers, a rare policy intervention generating sustained lift for low-income students.[10]

II. The Structure of Teacher Compensation

In the U.S., the median teacher receives $58,000 in annual salary and receives another $28,000 in benefits, primarily through health insurance and retirement.[11] The National Compensation Survey (NCS) reports that the costs of employing American primary and secondary school teachers are divided 69 percent toward salary, 11 percent toward health benefits, and 11 percent toward retirement benefits. The remaining 9 percent of compensation costs constitute legally required benefits, other pay (usually comprising bonuses), and paid leave.[12] Though typical civilian workers earn a slightly larger fraction of their compensation in salary, the primary difference in the structure of teacher pay is in benefits. Most significantly, other workers earn 20 percent less of their income in health benefits, half as much in retirement, and earn an order of magnitude *more* in supplemental pay, largely reflecting the fact that few schools employ reward-based compensation (Figlio and Kenny 2007; Bureau of Labor Statistics, 2016).

I use data from the Local Education Finance Survey (LEFS) which collects financial information from each school district to understand the total level of compensation, in the target school district, Aldine Independent School District, and across the broader state. The Aldine school district spent $89,461 per teacher in 2014; these data reflect the fact that Texas schools pay a smaller fraction of their compensation in benefits (26.1 percent) and a larger fraction in salary (73.9 percent) when compared to other states. A Freedom of Information request (FOIA) to the school district reveals a similar picture with 74.1 percent of their pay received as salary and 25.9 percent received in

[10] See, for instance, Altonji and Mansfield (2011), Dahl, Kostol, and Mogstad 2014, Chetty, Friedman, and Rockoff 2014, Heckman, Humphries, and Veramendi forthcoming
[11] This tally does not include special retirement health plans schools provide or the underfunding of pensions that the government presents as guaranteed (Farmer 2014; Novy-Marx and Rauh 2014). Government contributions would have to rise by 24.1 percent of payroll (a more than doubling from its current contribution of 16.3 percent of payroll), as of a few years ago, to close the fiscal gap on its retirement promises,
[12] The parallel shares for a generic civilian worker are 68.7 percent in salary, 8.8 percent in health benefits, and 5.2 percent in retirement. https://www.bls.gov/news.release/archives/ecec_03102016.htm

benefits. The school district reports paying, on behalf of the average worker, \$62,186 in salary, \$3,960 toward health insurance, \$5,161 toward pension, \$964 for retirement healthcare, and \$0 in merit rewards.

These three data sources (NCS, LEFS, and the FOIA disclosure) understate the extent to which state and local agencies will spend on benefits since they do not reflect the total cost of pension and retirement health plans which are underfunded but essentially guaranteed (Novy-Marx and Rauh 2014). My calculations suggest, for instance, that the state would need to double its contribution to retired health benefits and triple its pension contribution to be able to deliver on its promises reliably, in the absence of which the government may be forced make up that contribution in the future.[13] In the base analysis when calculating optimal compensation structure under various criteria, I exclude these implied costs since it may politically infeasible, or otherwise unlikely, to spend that money today, for instance, on salaries, which is likely why these personnel costs are deferred to the future to begin with.

III. Experimental Design and Econometric Framework

The Empirical Challenge

When economists set out to estimate preferences, they typically collect data on the choices people make and the options available to them at the time of choosing. Unfortunately, the records needed construct choice sets from which teachers select are unavailable and would be insufficient for our purposes. In the dozens of districts I informally interviewed, none kept records of offers made, precluding the assembly of what offers a teacher had to select from. For years, I approached firms that provided 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

[13] Though the issue appears far from settled, several judges have rejected attempts made by local officials to reduce pension benefits in the future, and the BLS describes pension benefits as "guaranteed" (Reuters 2013; Vinicky 2013; BLS 2013)

error; essentially no one accepted their offer through the interface. Moreover, the job market is highly decentralized, and schools make offers at widely varying times; teachers I spoke to very 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.

Even if those data were attainable, however, they would not be particularly informative. It is common for competing schools to have identical compensation structures, tenure timetables, and rules governing working conditions like class size. Across districts even, variation is extremely limited by statewide requirements and the common influence of the union bargaining. Districts within a state will often share a common pension programs, health-insurance plan, salary schedule, and class-size regulations. Where variation may exist at the borders between districts, the wealthier district usually offers a work setting which dominates the neighboring district in every dimension, providing no information on preferences whatever other than that which was already known: that more is preferred.[14] Choices along the borders of neighboring states suffer similar problems while being complicated by the fixed cost teachers face when acquiring a teaching certificate in a second jurisdiction.

Omitted variables present an additional layer of difficulty. Variation in pay is expected to be highly correlated with other factors (e.g., quality of facilities, staff competence, principal leadership, student characteristics, commuting time, etc.), making it difficult to separate the influence of compensation structure on teacher choice in realized data. The choice experiment provided here addresses the influence of other, unobserved factors to shed light on preferences by using a controlled setting in which there is no unobserved heterogeneity. Additionally, the controlled choice experiment allows teachers to express choice over work settings that do not occur at present in

[14] This empirical problem is endemic 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—are all visibly gerrymandered out of the largely poor Los Angeles Unified School District, and each area has its own distinct school district, some of which are the most highly rated in California and the U.S.

schools, like implementing merit rewards or transitioning to a defined-contribution retirement program. These are precisely the issues in teacher compensation that make the study of preferences challenging and, in some cases, impossible with naturally occurring data.

Choice Experiments and Conjoint Analysis

The conjoint survey is a tool developed to measure consumer preferences and forecast demand for components on a prospective product or service bundle, like cars in development or a potential software application (Beggs, Cardell, and Hausman 1981; Green and Srinivasan 1990; Hainmueller, Hopkins, and Yamamoto 2013). In parallel, this paper aims to estimate teacher utility over prospective compensation structures, contract terms, and working conditions. Doing so is particularly useful among public schools and other government employers since they function as a monopsony employer, and there is no natural market test (de Ree et al. 2017). 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 survey design. Each choice set requires the teacher to make a tradeoff of some kind and the package maximizes efficiency of the parameters of the utility model for a given number of choice sets.[15] The method allows the analyst to evaluate several hypotheses in a single study and compare the influence of various factors in teacher choice within the same setting, so that estimates are comparable. Moreover, the method is useful for avoiding social-desirability bias: in addition to being an essentially anonymous online survey, respondents have available multiple reasons to justify any particular choice in the conjoint setting (Hainmueller, Hopkins, and Yamamoto 2010).

In this survey I consider fourteen attributes recommended by the literature and interviews with experts.[16] These attributes include (1) starting salary, (2) salary growth rate, (3) health insurance

[15] To make the idea concrete, we never present a choice set which we know will be preferred since no option ever dominates its alternative in each of the attributes. Without this guarantee, we might present questions that illuminate very little.
[16] Many thanks to those experts willing to speak with me including Dale Ballou, Sean Corcoran, Susan Johnson, Thomas Kane, Rebecca Maynard, Michael Podgursky, and Jonah Rockoff.

plan, in terms of the deductible and monthly premia, (4) retirement income plan (replacement rate and defined benefits or defined contribution), (5) merit reward program, (6) class size, (7) the duration of the probationary contract, (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. Each attribute can take on several values, described in online Appendix Table 1.[17]

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. Thus, 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 deck to allow for comparisons between attributes in different decks. The second deck also includes starting salary (to "bridge" the decks), but also 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 time units, and how many hours of instructional support are provided the teacher each week. The third and final deck asks teachers to choose between job offers that vary starting salary (again, to assimilate the

[17] 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 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.

decks), rate of student poverty, student minority share, average achievement percentile, and whether a principal was “supportive” or “hands-off” with disruptive students. 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. Examples of these survey questions are presented in online Appendix Figures 1–3.

Because the survey is distributed by an organization hired to make recommendations regarding the district’s compensation structure, teachers have an incentive to thoughtfully consider and reveal their preferences, understanding that this information is collected to inform the new compensation structure. Teacher responses are completely confidential and have been reliably private in previous surveys implemented by the consulting group with whom I partnered; thus, teachers have little reason to believe their employer will ever be able to review their individual response. This setting is not formally strategy proof, but there’s good reason to believe that teachers responses are reflective of preferences. Hypothetical choice experiments, in a variety of settings, successfully predict individual choice behavior and willingness to pay in natural settings, even absent express incentive to reveal preferences truthfully (Hainmueller, Hangartner and Yamamoto 2015; Wlomert and Eggers 2016; Parker and Souleles 2017; Wiswall and Zafar 2017). [18] What’s more, formally incentive-compatible designs do not significantly alter the predictive validity of such experiments (Holt and Laury 2002; Ding 2007; Wlomert and Eggers 2016). Incentive compatibility seems to matter only if responding requires significant effort, or if subjects have a distinct reason to misrepresent their preferences (Camerer and Hogarth 1999).[19] Because the compensation and

[18] Three-quarters of the time, one’s conjoint responses correctly predicts your market behavior Wlomert and Eggers (2016). Similar predictive ability is seen in Brazell et al. (2006) and Iyengar and Jedidi (2012) .

[19] 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… In other tasks, incentives can actually hurt, if increased incentives cause people to overlearn a heuristic…, to overreact to feedback…to exert “too much effort” when a low-effort habit would suffice… or when arousal caused by incentives raises self-consciousness…In a few tasks, incentives appear to actually hurt. All of these are judgement or decision tasks.”

working conditions are a large part of a teacher's life and waking hours, teachers have likely considered their preferences, suggesting the need for new effort is minimal, a feature conducive to truth-telling. An older literature, too, finds that conjoint responses are strongly predictive of an individual's later choices (Robinson 1980; Srinivasan 1988) and out-of-sample market share (Benbenisty 1983; Clarke 1987).

To evaluate whether revealed preferences are rational, I test whether choice is monotonic with respect to ordered variables with clear correspondence to utility (Hainmueller and Hiscox 2010). I find that choosing an offer is strongly increasing in starting salary, salary growth, retirement replacement rate, class size reductions, and support provided to teachers, with teachers significantly more likely to select the highest categories than the medium one, and significantly more likely to select a medium category than the lowest, a result that holds when making within-teacher comparisons.

It could be that by asking teachers to make tradeoffs between hypothetical job offers we are implicitly asking them to value things they may not care about in real choice settings. To address this concern, I include in the choice sets an irrelevant school feature—whether the school bus at the featured school is blue (McFadden 1981)—to evaluate whether teachers demand things that have no impact on a teacher's welfare, and I find that teachers reliably express no preference over this irrelevant detail, aiding a causal interpretation. Uninstructed, subjects may fill in the state space, inferring other characteristics that influence their preference other than those features explicitly described. To address this concern, 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 generates classical measurement error in the outcome variable, choice, which does not bias the results but reduces precision. Second, the survey was 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 an unspecified amount of time. Third, the online survey environment records how long each teacher takes to respond to each question, and teachers appear to take sufficient 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, except in the estimated preference for defined-contributions plan.[20] Since respondents taking more time to express effectively the same preferences, this suggests that more attention isn't necessary to elicit more accurate preferences. This also suggests that a subset of teachers do not resort to simplifying rules by paying attention to a few attributes and not others. If that were the case, we would expect measured preferences to be distinct for subjects expending more time and effort to consider each question.

I deployed the experiment in Aldine ISD, 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 responded to the survey. The high response rate was encouraged by school-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 job offers, where each selection forces the teacher to make a tradeoff between two or more features that are assumed to generate positive utility. For instance, one option may provide a more generous salary, but comes at the cost of a larger class; or, a more generous retirement plan accompanies a smaller potential for merit pay. Under weak conditions, the hypothetical job selection data identify

[20] To identify people who take longer, I regress response time on question and teacher dummy variables. The composite of the residual plus the teacher fixed effect reflects the average residualized time that the teacher took to respond to questions. I implement this procedure rather than just using responses that took more or less time in case some questions were harder to compute than others. The method used identifies people who systematically take longer and shorter durations of time when rendering a decision. 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 ($p<0.001$).

job preferences over several factors while standard realized choice data do not (Wiswall and Zafar 2017). 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. I assume the utility from offer x is additively separable.

Offers are indexed by n, and there are is a finite set of offers $j = 1, \ldots, J$. 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 like class size and time to tenure. To explore the influence of each factor, I use a linear-probability model, regressing respondent choices on a vector of characteristics, conditioning on option fixed effects to account for the choices available to the teacher:

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

Here, teacher i prefers option j from choice set S. For comparison, I also present the results from conditional logistic regression (Louviere et al. 2000). In each, part-worth utilities are denoted β and characteristics of alternative n are given by X_n. In the main analysis, the linear-probability model is marginally more successful in explaining choice variation and in accurately predicting the choices of subjects. For example, the LPM accurately predicts 64 percent of choices whereas the conditional logit predicts slightly fewer, at 62 percent, in the working-condition deck. Throughout the empirical work, the standard errors are clustered by teacher ID.

III. Results

Teacher Utility over Compensation and Working Conditions

The main results are presented in Figures 1–3 and Table 2. The figures visualize the results nonparametrically by estimating the model with bins, providing a transparent way to evaluate the results and compare the influence of different attributes. In the Table, 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 whereas the right three represent estimates from the conditional logistic model estimated with maximum likelihood. All estimates are harmonized across the three decks using subjects' response to the salary feature in each deck.[21] Columns (3) and (6) represent a money metric which measures how much teachers value a unit of that feature in terms of a salary increase.

Teachers value $1,000 of district subsidies for insurance equal to $802 in salary increases, suggesting that schools may present a more attractive compensation package by allocating these resources to salary.[22] Merit pay of $1,000 for the top quarter of teachers in terms of principal evaluations and value-added measures is valued $318, a quarter more than its cost. Moving to a defined-contribution (DC) retirement plans from a traditional pension increases teacher utility equal to a starting-salary increase of $851, presumably because DC plans are portable and less subject to political risk; prior work finds that public workers are concerned about the underfunding of pensions (Ehrenberg 1980; Inman 1980; Smith 1981). Teachers value an additional ten-point replacement rate at retirement equivalent as a $1,520 salary increase, conditional on retirement plan. An additional one-percent increase in salary growth is valued equivalent to a $2,093 increase in starting salary increase.

The options varied how many years teachers were evaluated before being provided 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 $415 salary increase. The district also has regular review periods in which the teacher's performance is reviewed once a teacher has permanent status. More frequent reviews impose little disutility. An hour commute is

[21] 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 relative units. Each coefficient in Deck 3 is multiplied by $\beta_{salary}^{Deck1}/\beta_{salary}^{Deck3}$. The salary betas are similar across the decks, but differ slightly. This is not an issue because, in preference estimation, only relative parameters matter (Train 2009).

[22] This is especially interesting because salary faces marginal income tax rates as well as payroll taxes where the insurance subsidy does not.

equivalent to a salary reduction of $2,395, suggesting that teachers are willing to be paid $6.70 per hour to commute to work, less than the minimum wage. This is consistent with work reporting that people derive utility from commuting or fail to fully assess the costs of commuting, or actually prefer travel time to work (Richter 1990; Redmond and Mokhtarian 2001; Stutzer and Frey 2008; Frey and Stutzer 2014).[23] Two papers in the study of teacher labor markets have also estimated how commuting affects teacher preferences, using structural models which suggest larger disutility from commuting.[24] A class-size reduction of one student increases teacher utility the equivalent to a $490 salary increase (1 percent of average salary). Although I am not aware of previous studies that directly estimate WTP for reductions in class size, WTP can be calculated from estimates of the influence of class size on teacher retention. The estimates I recover are in the middle of the available evidence. 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), in contrast, finds no significant relationship between class size and teacher turnover. Teachers value an additional hour of teaching assistance each week at $195, less than the cost of hiring someone to provide assistance at minimum wage, potentially consistent with the costly nature of transferring tasks (Goldin 2014).

The final deck varied student attributes and school-leadership characteristics. Teachers prefer schools with higher-achieving students and fewer children in poverty, but teachers have no preference over the racial composition of their students, surprisingly similar to the results of Antos and Rosen (1975) who finds no racial preference among white teachers but preferences for schools housing children with higher achievement and less poverty; this result may help interpret findings that teacher quality declines in response to segregation, as in Jackson (2009) and Cook (2018). A ten

[23] Analysts report that workers value non-zero commute times as a mental transition from work to home, as a scenic or relaxing routine, or to use the time for recreation, including listening to music, radio, podcasts, or books.

[24] Biasi (2017) finds that teachers in Wisconsin are willing to pay 0.4 percent of salary to be 1 percent closer to work. Using the choice experiment, I find, in some contrast, that a teacher would be willing to pay fewer than 0.1 percent of salary for an equivalent increase in proximity. Boyd et al. (2013) also use a model to estimate the cost of commuting. It is difficult to compare their estimates to mine because they cannot generate precise measures of utility for salary, which makes for the natural standard of comparison. They find that a one-percent increase in distance is equivalent to a 1.5 percent increase in the share of students in poverty; again suggesting a disutility of commuting larger than those I recover in the experimental setting.

percentage-point reduction in student poverty is equivalent to a salary increase of $343 (0.6 percent of the average salary in Aldine). Prior analysts have noted that the psychic costs of teaching low-income students lead higher quality teachers to leave low-income schools, yielding an obstacle for equal opportunity without compensating differentials (for example, Lankford, Loeb, and Wyckoff 2002; Mansfield 2015). How much would it cost for Aldine to compensates teachers for teaching in schools with many low-income students? In Aldine, it would cost $8.2 million each year (about 3 percent of the budget Aldine now spends on teacher compensation) to equalize teacher utility to the level of a typical suburban school.[25] Because Aldine is largely low income, the teachers' preferences are those of a highly selected sample. The needed compensating differential for the marginal teacher, outside the district, is likely higher. Student achievement is important to teachers: a ten-point increase in the average percentile at which students perform is worth $510.

The most predictive attribute in any deck is whether the principal is "supportive" or "hands-off" of teachers with disruptive students.[26] Having a supportive principal provides utility equivalent to $8,567 increase in starting 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 setting.[27] This finding develops reporting that student discipline and behavior problems are costly and lead some teachers to depart the profession (Ingersoll 2003; Kinsler 2013; Luczak, Darling-Hammond, and Loeb 2013; Carrell, Hoekstra, and Kuka 2018). A related question is whether supportive principals are more important in difficult school settings. I estimate linear probability models with binary measures of whether a school is low-achieving or high-poverty are interacted

[25] Aldine has twice the share of low-income students as a typical suburban district in the U.S. This is calculated by multiplying the difference in low-income share between Aldine and the average suburban school (88.6 – 45.9 = 42.7), and multiplying that by the number of teachers in Aldine, and then multiplied by the salary value of a 1 percent low-income share change to teachers. Aldine ISD may already have increased teacher compensation to address this need; in that case, this estimate would suggest at most what share of compensation is for equalizing socioeconomic utility to teachers. It remains a surprising feature of the education system that teachers are not expressly compensated for the difficulty of the circumstances in which they teach. Extrapolating this estimate, to raise the utility of all teachers in school-wide Title-I districts nationwide to the demographic utility of a typical suburban school, it would cost $1.6 billion annually.

[26] I contacted a random sample of respondents to ask them what they imagined a supportive principal doing...

[27] To understand what teachers understood by having a supportive or hands-off principal with disruptive students, I emailed a number of respondents. They tended to say that a supportive principal would talk with disruptive students, support the teacher's decisions rules, and engage with teachers to find solutions, including partnering the teacher with adults at the school who had built a rapport with the disruptive student.

with the supportive-principal indicator. Supportive principals completely erase the utility loss of working in a low-achieving school, and they mitigate 59 percent of the disutility associated with teaching in a high-poverty school (Table 2).

Scope for Separating Equilibria

Whether or not compensation and working conditions can generate a separating equilibrium in which better teachers differentially select into and then remain in a school depends on whether excellent teachers have distinct preferences from their colleagues. Moreover, it may be important to see if the best teachers have different preferences for working conditions that are not affected by policy, like student characteristics, to understand whether larger compensating differentials are needed to draw high-quality teachers into needy schools.

To evaluate teacher quality, I estimate value-add models from student data and incorporate Danielson observation scores. The student data contain test scores for each student in each year they are tested and linked to each teacher a student was assigned covering 2011–2016. I estimate VAMs using all the available test scores that a student has from their previous school year while controlling for whether a test score is missing. The VAM used in the primary analysis is the average of the subject-specific VAMs available, usually math and reading. The resulting VAMs are 0.1 on average with a standard deviation of 1. I translate teacher VAMs into ten deciles and generate a quality index from those deciles from 0 to 1. There are records sufficient to estimate value-add for just under half of teachers (44 percent), since students are not tested in all grades. To provide a measure of quality that covers a broader array of teachers, I incorporate Danielson observation scores for teachers without VAMs.

The district evaluates its teachers using a Danielson rubric in which the principal rates its teacher in four categories: “planning and preparation,” “classroom environment,” “instruction,” and “professional responsibilities” on a scale from 1 (ineffective) to 4 (highly effective). I sum each teacher's four scores and assign deciles based on the total score and generate a parallel quality index from 0 to 1. The VAM index and the Danielson index are highly correlated for teachers with both measures ($p < 0.001$), and one predicts about 5 percent of the other. For those teachers who do not have a VAM index, I input the Danielson index as their quality measure, what I call "the quality index." Together, the VAM index and the Danielson index provide a quality measure for about 80 percent of respondents, though the same results present themselves (with less statistical power) when each measure is used in isolation.

I interact each of the factors from Table 2 with the quality index in a model of teacher choice. I also present visual evidence by interacting quality-decile dummies with each trait and plotting the interactions to visualize how teachers of varying quality respond to each attribute. To account for the fact that more experienced teachers may systematically have higher value-add and have distinct preferences for related to experience and not their underlying ability, I condition on experience dummies that indicate having exactly n years of experience in the table and visual results.

The best and worst teachers have quite similar preferences in general (Table 3, online Appendix Tables 3 and 4). Great teachers do not, for instance, have a stronger preferences for high-performing or wealthier 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, and retirement benefits. The only way in which high-performing teachers systematically differ in my data is their preferences for offers including merit rewards (see Table 3 and Figure 4). A teacher in the bottom decile values a $1,000 merit reward as the equivalent to a $172 salary increase. Teachers in the top decile value the same merit program as the equivalent to a $581 salary increase (the interaction $p < 0.001$). If teachers received two offers with the same expected value, the high-quality (top decile) type is 21 percent more likely than a bottom-decile teacher to select an offer providing $3,000 in merit pay per year. Over time, this could generate meaningful positive selection in attrition, or, under certain conditions, in recruiting. Since they value the increased compensation or recognition, their utility is higher, and the probability of attrition is reduced relative to low-quality teachers.

The decile plots demonstrate the same story. Figure 4 demonstrates how teachers of each quality decile express preferences for an increment of merit pay relative to bottom-decile teachers. Deciles 2 through 7 express differential preferences that are very close to zero. Teachers in deciles 8 through 10, however, each have significantly stronger preferences for merit pay than low-decile teachers. The top decile is 3.3 percent ($p < 0.05$) more likely to select an offer providing $1,000 in merit pay, and a teacher in the next top decile is 4.2 percent ($p < 0.001$) more likely. I present the corollary plot for each of the other school attributes in online Appendix figures 4 – 6, each lacking a systematic pattern, which is consistent with the results in Table 3 and in online Appendix Tables 5 and 6 in which higher quality teachers do not differ significantly in their preferences for school attributes.

Teacher-Utility Optimal Compensation Structure

I simulate the optimal structure of teacher compensation under two related objective functions: First, we consider teacher-utility maximization, and then we turn to student-achievement maximization which embeds the teacher's utility as a factor in retention. Teacher-utility maximization may be the attitude of districts with strong unions that aggregate and represent the preferences of their members (Farber 1978). By understanding the teacher-optimal structure, schools can improve the well-being of their teachers by reallocating scarce district resources, requiring no new financial contributions. 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 each non-binary feature (online Appendix Table XXX) which blends utility estimates on compensation from Deck 1 and utility estimates on class size from Deck 2. To determine what is optimal, I specify costs for the budget constraint. The function accounts for the costs of starting salary, the rate of salary growth, retirement replacement, guaranteed pensions, merit pay, and the cost of recruiting and training more often, a function of attrition. The costs interact, for instance, since retirement replacement becomes more expensive as base pay or salary growth increase. Class-size reductions also become costlier as total compensation rises since it becomes increasingly expensive to hire an additional teacher to shrink class size. The details of the cost function can be found in online Appendix 1.

To search for optimal locations in compensation space, I generate a large grid in which each compensation level is independent and varies uniformly along a large range. I iteratively calculate the utility and cost of each bundle, and regenerate random compensation structure values for bundles that do not meet the budget constraint (either being more or less than the budget constraint in a small range), a process I repeat a hundred times. Having generated several feasible bundles from random draws, I next identify the range of compensation values that achieve high-value, feasible bundles and iteratively generate more bundles within that range. Specifically, I evaluate the range of each compensation value among bundles that equal the budget constraint and generate utility above the 75^{th} percentile of the simulation; I generate additional bundle vectors from the narrower range of apparently successful values and iterate this process a thousand times. Then I average the vector values of the top 0.01 percent of simulations.

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 merit rewards, had an average class size of 28.7 students, paid $3,960 in health-insurance subsidies, and promised to replace 69 percent of your top earnings in retirement through a guaranteed pension program if the teacher remained for 30 years. To maximize teacher utility subject to the same budget the school currently uses on personnel, the school would pay 10 percent more in base salary ($55,000) and offer $1,700 in merit pay to the top quartile. These increases are financed by reduced expenditure elsewhere: Increased class size (3 percent), reductions in salary growth (28 percent), reduced health insurance subsidies (100 percent), a transition to defined contribution plans with a reduced replacement rate (13 percent). I also solve the optimization problem using a nonlinear programming solver and find an essentially identical solution.

In total, these changes cost the district no more but increase teacher welfare by 2.2 percent, the equivalent of a $3,400 increase in salary. In the next section, I discuss the creation of a student-achievement production function. The teacher-optimal compensation structure would increase student achievement by 0.06σ over the long-term.

Student-Achievement Optimal

Schools are tasked with aims other than maximizing teacher utility, and teacher welfare does not directly increase student achievement (De Ree et al. 2017). A policymaker may instead spur schools to promote achievement to the greatest extent possible with available resources. I specify an achievement production function using averages of domestic estimates or, when available, recent estimates from Texas; in the achievement function, students learn somewhat more in smaller classes (Krueger 1999; Hoxby 2000), somewhat more under merit rewards (Lavy 2009; Imberman and Lovenheim 2015; Bond and Mumford 2018), merit compensation induces selection into the school based on teacher quality (Biasi 2018), and teacher utility affects the distribution of experience (Hendricks 2014), with more experienced teachers having increasing, concave impacts on students. I implement the same simulation process as described above, but instead calculate achievement as the objective function of interest. The details of this functional specification are discussed at greater length in online Appendix XXX.

In comparison to how the district now compensates teachers, the structure that maximizes achievement would pay teachers 2 percent less in base pay, nearly double in salary growth (3.5 percent growth rate), $4,000 in merit pay, and a class size that's 5 percent larger. The reason for the more modest starting pay with greater income growth is that it focuses the retention effort on teachers who already have experience. Like the utility-maximizing bundle, the achievement-maximizing bundle reduces the district subsidy for health insurance in favor of other compensation vehicles. The achievement-optimal bundle also reduces retirement generosity by 20 percent while moving to a defined-contributions retirement plan. This structure increases teacher utility by about 0.2%, the equivalent of a $231 salary increase, but the structure provides 0.23σ in terms of student achievement in the long-term. The achievement gains come in least measure from effort induced by merit pay (20 percent), then increased teacher retention and experience (26 percent), and from differential selection and attrition induced by merit pay (54 percent).

These estimates are calculated based on partial-equilibrium framework in which one school, say a charter, adopts optimal compensation, in a sea of schools who provide an equally costly, but less valuable, compensation structure. The achievement gains are fully realized in the long term by affecting the selection and retention of teachers over many years. That is, with the exception of induced effort, selection and retention impacts the cohorts exposed to optimal compensation, though merit pay may induce differential retention in the short term as well (Hoxby and Leigh 2004; Biasi 2017). One question of interest is whether merit pay can induce a positive selection into teaching if broadly adopted. Though the question is beyond the reach of this empirical setting, two important conditions are necessary for merit pay to induce positive selection. First, prospective teachers would need to have a signal, before they embark in the profession, whether they would be teachers who would benefit from merit pay. If the beliefs of prospective teachers about their self-efficacy is uncorrelated to their eventual quality, merit pay programs would fail to drive positive selection. Second, marginal teachers, those who could be induced into teaching, must have similar, affirmative preferences for merit pay as other teachers.

To a limited degree, the setting allows weak tests of each of these. First, teachers in Aldine are never provided their value-add score, though they do receive scores from the central office based on their Danielson evaluations. If teachers have private information, high-VAM teachers would prefer merit pay, even after conditioning on their known observation score. Consistent with private information about VAMs, I find that high-VAM teachers have stronger preferences for merit pay, even conditioning on their Danielson scores. This is not strong evidence, however, since teacher may learn by experience teaching whether they have a strong impact on student achievement, in part if they intuit their VAM from the test score trajectories of their students. In Aldine, teachers do not receive VAM scores, but they do receive a report of their students' test scores before the school year begins, and again for those same students after the year concludes. Though they are never given both sets of scores simultaneously, a careful teacher that maintained her records may be able to infer her VAM. Recall that the merit-pay scheme in this experiment awards a bonus to teachers in the top 25 percent of a school. Because of this, teachers would also need to know the distribution of VAM scores for other teachers to make a judgement about where they fall in the distribution of teachers.

A comparison of the teacher-optimal and student-optimal compensation structures is instructive. First, maximizing either objective function would increase district spending on salary. I simulate a retention model based on Hendricks (2013) in which teachers are more likely to be retained under higher compensation, but the sensitivity of attrition to compensation attenuates over the life cycle. I simulate the average salary compensation in the status-quo regime to be an average expenditure of $58,460 per teacher per year. Under utility-optimal compensation, the district spends ten percent more in base pay but increases compensation for experience at a lower rate. Accounting for retention, which declines somewhat, this increases the average expense on salary to $61,130 (5 percent). Notice the retention declines, despite spending more per teacher. This is because this salary increases broadly raise the pay of inexperienced teachers increasing the pay of inexperienced teachers whose retention is less important to the production of achievement. The student-optimal bundle spends 18 percent more on salary but increases total retention significantly (the average years of experience increases 12 percent).

There is, moreover, an increase in district spending on merit compensation, which is relatively inexpensive since it applies to a fraction of teachers each year. These increases are funded by cost reductions in terms of reduced insurance subsidies, slightly larger class sizes, and a less generous retirement plan, which are less valued by teachers or have a smaller direct effect on achievement.

V. Discussion and Conclusion

There were more attributes to test than there was space. At the end of the survey, I asked teachers to evaluate several attributes by guessing how likely they would be to accept offers as I varied a single attribute. For instance, we find in the choice data that teachers are enormously responsive to principal support with disruptive students. In the baseline case, the offer features a principal who is "mostly involved in curriculum and school management," which teachers say they would be 69.5% likely to take the job. Teachers, however, are more likely to accept an offer (85.9%) in which the principal "will be supportive in disciplining disruptive students," consistent with the results presented previous. I find that offers featuring a principal who "will support the teacher in a disagreement with a parent" is nearly as important as support with disruptive students; teachers report a 82.8% odds of accepting this offer.

Another question not answered in the discrete-choice analysis above is whether teachers prefer being evaluated by value-added measures (VAM) or by observation scores. One could imagine that teachers prefer Danielson scores because they are personal, and also because they can be managed less costlessly. While a teacher can prepare for a scheduled observation, success in the VAM is proportional to her skills and effort. On the other hand, teachers may prefer an objective measure to an observation score which could be used politically by principals to privilege their friends. The conjoint analysis provided inconsistent evidence on this question. Somewhat surprisingly, teachers preferred a merit reward based on VAM-only (58.2%); next most preferred was merit reward based on VAM and observation scores (55.9%); and least preferred of all was to be evaluated by observation score alone (51.1%)

Interestingly, the district's compensation scheme does not conform either to teacher preferences or achievement maximization. Since unions are led by older, veteran teachers, it may be that they bargain for compensation structures that benefit themselves more so than the profession broadly.[28] I explore this explanation with some suggestive evidence using school-finance data linked to union-strength measures for each state. If preferences of union representatives explain district reliance on health and retirement benefits over salary, we might expect places with stronger union presence to pay a larger share of compensation in benefits, conditional on total compensation.[29] I gather a state-level union-strength measure 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 generate 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 2.5 percentage points ($p < 0.001$), explaining a ten-point difference between states with the weakest unions (where compensation is 28 percent benefits) and where unions are strongest (where compensation is 38 percent benefits), conditional on total compensation.

To assess the generalizability of the recommendations for optimal structure, I compare where Aldine ISD compensation structure to the rest of the state and country.[30] One of the consistent lessons from the optimization results is that Aldine may improve teacher welfare and student achievement by increasing salary expenditures as a fraction of total compensation. If Aldine has low salary share compared to other districts, Aldine may simply fall on the high side of a distribution centered on what is optimal. In online Appendix Figure 10, I show where Aldine's compensation

[28] Indeed, I find that teachers value more generous retirement plans the more senior they become, and the relationship is strictly monotonic for bins of teacher experience.
[29] 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. We control for total compensation to avoid confounding benefit-share increases with increased total compensation.
[30] Compared to teachers in other districts, teachers in Aldine ISD receive total compensations at the 55th percentile in Texas and the 24th percentile across the country. See, for reference, online Appendix Figure X.

falls in the distribution of US districts in terms of salary share. Two thirds of school districts have salary shares *below* Aldine; 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 Aldine. What this suggests is that, since Aldine underinvests in salary, the many school districts who invest even less are likely also underinvesting.

REFERENCES

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

Avery, Christopher N., Mark E. Glickman, Caroline M. Hoxby, and Andrew Metrick. "A Revealed Preference Ranking of US Colleges and Universities." *The Quarterly Journal of Economics* 128, no. 1 (2012): 425-467.

Bacolod, Marigee P. "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality." *The Review of Economics and Statistics* 89, no. 4 (2007): 737-751.

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

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

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

Benbenisty, Rochelle L. "Attitude Research, Conjoint Analysis Guided Ma Bell's Entry into Data Terminal Market." *Marketing News* 13 (1983): 12.

Biasi, Barbara. "Unions, Salaries, and the Market for Teachers: Evidence from Wisconsin." *Working paper* (2018).

Bond, Timothy, and Kevin J. Mumford. "Teacher Performance Pay in the United States: Incidence and Adult Outcomes." *Working Paper* (2018).

Boyd, Donald, Grossman, P., Lankford, H., Loeb, S., & Wyckoff, James. (2008). "Who Leaves? Teacher Attrition and Student Achievement." National Bureau of Economic Research, working paper N. 14022.

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.

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.

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.

Chetty, Raj, Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *The Quarterly Journal of Economics*, *126*(4), 1593-1660.

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

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

Chingos, Matthew M. (2012). "The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida's Statewide Mandate." *Economics of Education Review*, 31, 543–562.

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.

Cook, Jason. "Segregation, Student Achievement, and Postsecondary Attainment: Evidence from the Introduction of Race-Blind Magnet School Lotteries." (2016).

Corcoran, Sean P., William N. Evans, and Robert M. Schwab. "Changing Labor-Market Opportunities for Women and the Quality of Teachers, 1957-2000." *American Economic Review* 94, no. 2 (2004): 230-235.

Cowen, Tyler. *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*: A Penguin eSpecial from Dutton. Penguin, 2011.

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 (2017): 993-1039.

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.

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

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

Feng, Li, & Sass, Tim R. (2017). “Teacher Quality and Teacher Mobility.” *Education Finance and Policy*, *12*(3), 396-418.

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

Goldin, Claudia. "A grand gender convergence: Its last chapter." *American Economic Review* 104, no. 4 (2014): 1091-1119.

Gordon, Robert J. *The Rise and Fall of American Growth: The US Standard of Living since the Civil War*. Princeton University Press, 2016.

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 (2013): 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, E. A., & Rivkin, S. G. (2010). “Generalizations about Using Value-Added Measures of Teacher Quality.” *The American Economic Review*, *100*(2), 267-271.

Harris, Douglas N., and Tim R. Sass. “Teacher Training, Teacher Quality and Student Achievement.” *Journal of Public Economics* 95, no. 7-8 (2011): 798-812.

Hendricks, Matthew D. “Does It Pay to Pay Teachers More? Evidence from Texas.” *Journal of Public Economics* 109 (2014): 50-63.

Hess, Richard (2010). “Missing the point: Tomorrow’s Big Merit Pay Study Will Tell Us…Nothing.” *Education Week,* September.

Holt, Charles A., and Susan K. Laury. “Risk Aversion and Incentive Effects.” *American Economic Review* 92, no. 5 (2002): 1644-1655.

Hoxby, Caroline M., & Leigh, Andrew. (2004). “Pulled away or Pushed out? Explaining the Decline of Teacher Aptitude in the United States.” *The American Economic Review*, *94*(2), 236-240.

Imazeki, J. (2005). “Teacher Salaries and Teacher Attrition.” *Economics of Education Review*, *24*(4), 431-449.

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.

Jensen, Robert. (2010). “The (Perceived) Returns to Education and the Demand for Schooling.” *The Quarterly Journal of Economics*, *125*(2), 515-548.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. “What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City.” *Economics of Education review* 27, no. 6 (2008): 615-631.

Kane, Thomas J., & Staiger, Douglas O. (2008). ”Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation” (No. w14607). National Bureau of Economic Research.

Kinsler, Josh. “School discipline: A Source or Salve for the Racial Achievement Gap?” *International Economic Review* 54, no. 1 (2013): 355-383.

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. “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.

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.

Luczak, John, Linda Darling-Hammond, and Susanna Loeb. "How Teaching Conditions Predict: Teacher Turnover in California Schools." In *Rendering School Resources More Effective*, pp. 48-99. Routledge, 2013.

Mansfield, Richard K. “Teacher Quality and Student Inequality.” *Journal of Labor Economics* 33, no. 3 (2015): 751-788.

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.

Murnane, Richard J., and Randall J. Olsen. “The Effect of Salaries and Opportunity Costs on Duration in Teaching: Evidence from Michigan.” *The Review of Economics and Statistics* (1989): 347-352.

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.

Parker, Jonathan A., and Nicholas S. Souleles. “Reported Preference vs. Revealed Preference: Evidence from the Propensity to Spend Tax Rebates.” No. w23920. National Bureau of Economic Research, 2017.

Podgursky, M. (2006). “Teams versus Bureaucracies: Personnel Policy, Wage-Setting, and Teacher Quality in Traditional Public, Charter, and Private Schools.” *Education and Policy Analysis*.

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

Robinson, Patrick J. “Applications of Conjoint Analysis to Pricing Problems.” In *Market Measurement and Analysis: Proceedings of the 1979 ORSA/TIMS Conference on Marketing*, pp. 183-205. Cambridge, MA: Marketing Science Institute, 1980.

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, J. (2010). “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement.” *The Quarterly Journal of Economics*, *125*(1), 175-214.

Rothstein, J. (2015). “Teacher Quality Policy When Supply Matters.” *American Economic Review*, *105*(1), 100-130.

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

Srinivasan, Venkat. “A Conjunctive-Compensatory Approach to the Self-Explication of Multi-attributed Preferences.” *Decision Sciences* 19, no. 2 (1988): 295-305.

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

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

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

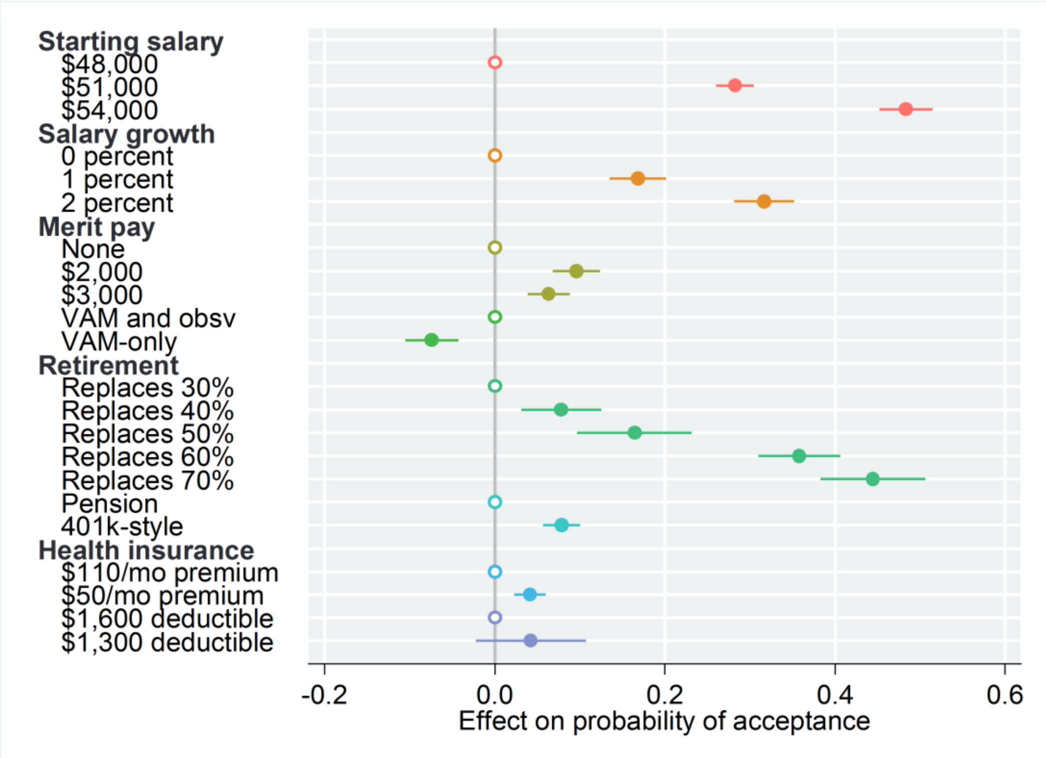
Wiswall, Matthew, and Basit Zafar. “Preference for the Workplace, Investment in Human Capital, and Gender.” *The Quarterly Journal of Economics* 133, no. 1 (2017): 457-507.

Wlömert, Nils, and Felix Eggers. “Predicting New Service Adoption with Conjoint Analysis: External Validity of BDM-Based Incentive-Aligned and Dual-Response Choice Designs.” *Marketing Letters* 27, no. 1 (2016): 195-210.

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

FIGURE 1—EFFECTS OF COMPENSATION ATTRIBUTES

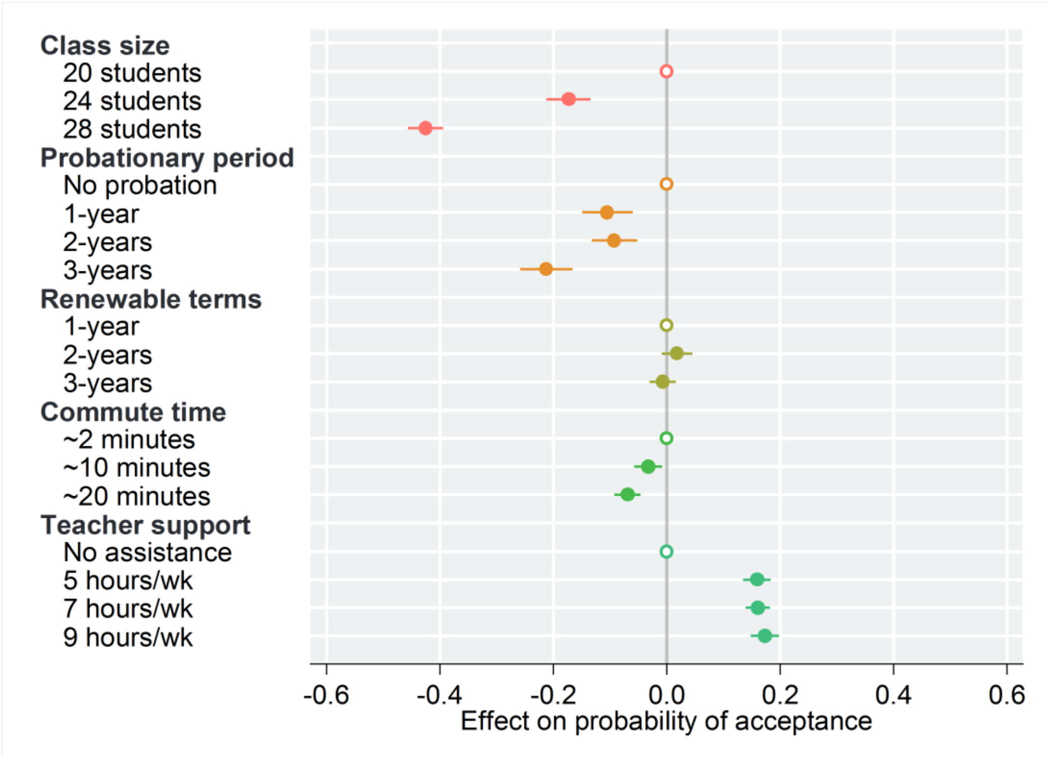
ON THE PROBABILITY THAT TEACHERS ACCEPT THE JOB OFFER



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.

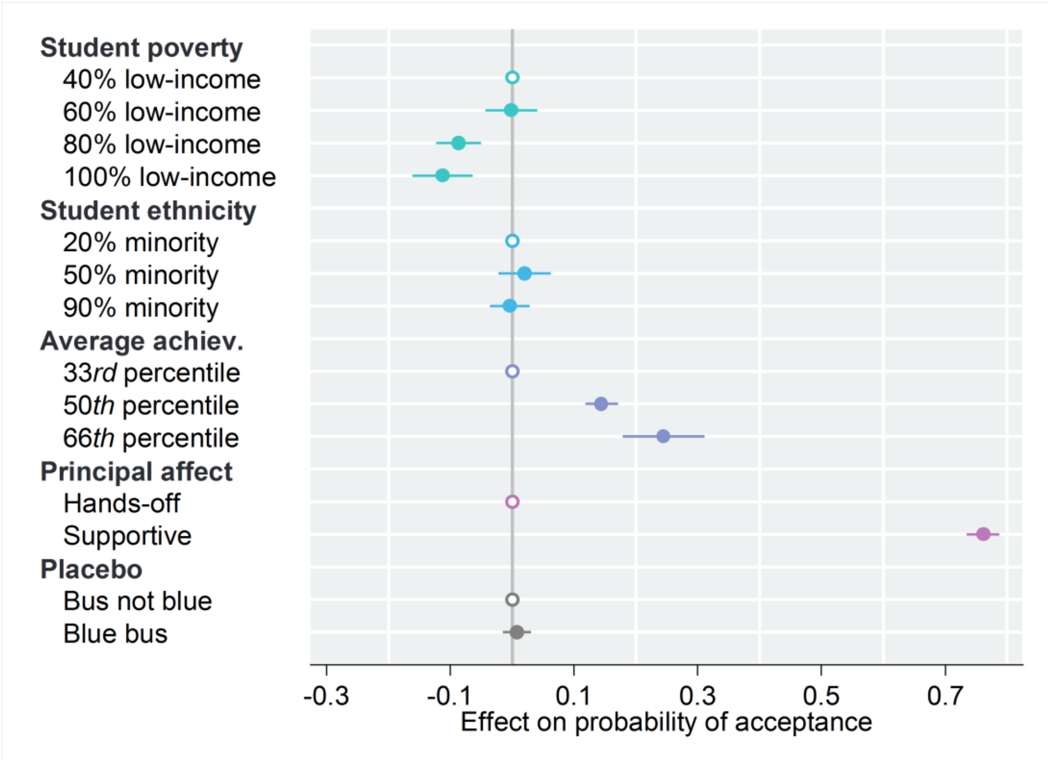
FIGURE 2—EFFECTS OF WORKING-CONDITION ATTRIBUTES

ON THE PROBABILITY THAT TEACHERS ACCEPT THE 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.

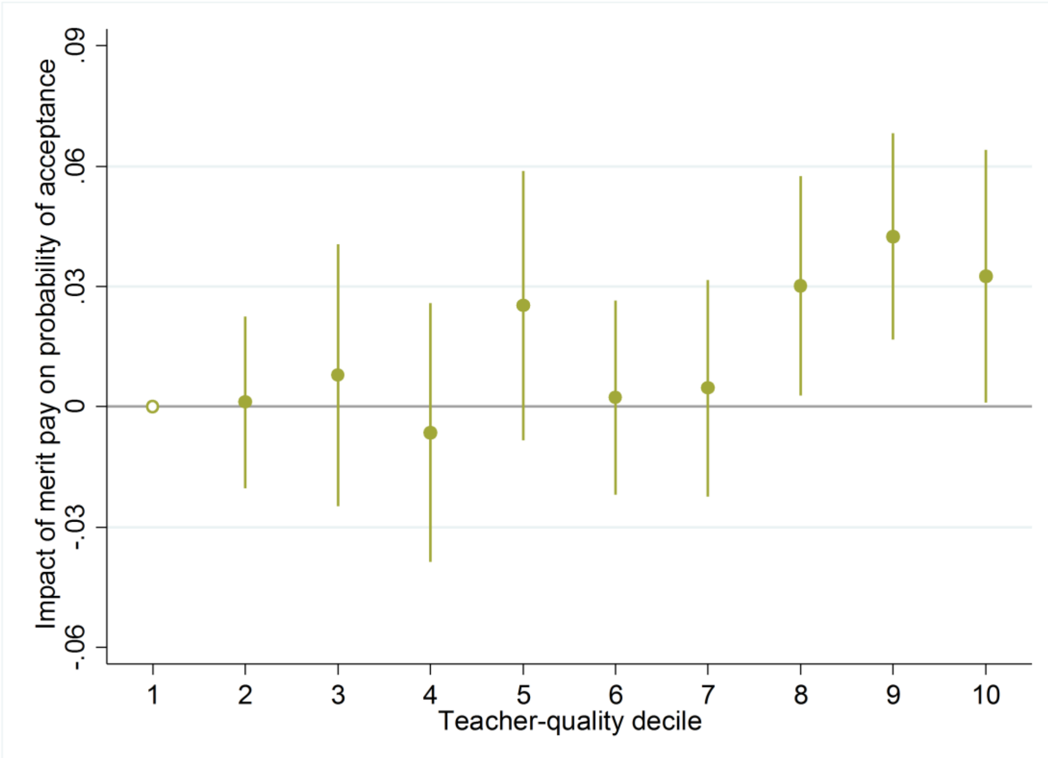
FIGURE 3—EFFECTS OF STUDENT AND PRINCIPAL ATTRIBUTES ON THE PROBABILITY THAT TEACHERS ACCEPT THE JOB OFFER



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 THE 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. The impacts plotted are in addition to the baseline impact of merit pay on accepting an offer, 1.6 percent (standard error: 0.007).

TABLE 1—SUMMARY STATISTICS ON OFFER ATTRIBUTES FOR CONJOINT EXPERIMENTS

	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 primary variables affecting choice in the 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	Relative	Coeff	SE	Relative
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: Compensation Deck						
Salary						
Starting salary	0.089**	(0.002)	$1,000	0.421**	(0.008)	$1,000
Salary growth	0.189**	(0.010)	$2,121	0.940**	(0.037)	$2,233
Merit reward						
Bonus amount	0.028**	(0.004)	$318	0.184**	(0.013)	$437
VAM only	-0.074**	(0.016)	-$829	-0.197**	(0.055)	-$468
Retirement						
Replacement	0.014**	(0.001)	$153	0.067**	(0.002)	$159
401k-style	0.076**	(0.011)	$855	0.418**	(0.035)	$993
Health insurance						
Premium (yearly)	-0.073**	(0.016)	-$815	-0.401**	(0.051)	-$952
Deductible	-0.442	(0.234)	-$4,961	-1.56	(0.811)	-$3,710
Panel 2: Working-Conditions Deck						
Contract						
Probationary period	-0.047**	(0.004)	-$415	-0.210**	(0.018)	-$308
Term length	-0.001	(0.004)	-$12	0.01	(0.017)	$13
Working conditions						
Commute time	-0.272**	(0.037)	-$2,395	-1.691**	(0.163)	-$2,486
Class size	-0.055**	(0.001)	-$488	-0.252**	(0.006)	-$370
Assistance	0.022**	(0.001)	$195	0.106**	(0.004)	$155
Panel 3: Students-&-Leaders Deck						
Students						
Percent low income	-0.030**	(0.004)	-$343	-0.120**	(0.011)	-$286
Percent minority	0.002	(0.002)	$17	0.00	(0.007)	-$1
Ave. achievement	0.050**	(0.004)	$510	0.220**	(0.012)	$527
Principal affect						
Supportive	0.760**	(0.014)	$8,567	3.040**	(0.044)	$7,236
Placebo						
Blue bus	0.010	(0.012)	$80	0.03	(0.039)	$64

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=26,074, %Predicted=0.65, R-squared=0.20; Deck 2: N=25,870, %Predicted=0.64, R-squared=0.28; Deck 3: N=19,404, %Predicted=0.75, R-squared=0.36.

TABLE 3—TEACHER PREFERENCES BY QUALITY

	Choice	
	Reference Group (1)	Quality-index interaction (2)
Salary		
Starting salary	0.091**	-0.001
	(0.004)	(0.006)
Salary growth	0.173**	0.013
	(0.014)	(0.018)
Merit reward		
Bonus amount	0.015*	0.039**
	(0.007)	(0.011)
VAM only	-0.053*	-0.047
	(0.023)	(0.028)
Retirement		
Replacement	0.013**	0.002
	(0.001)	(0.001)
401k-style	0.068**	0.019
	(0.020)	(0.031)
Health insurance		
Premium (yearly)	-0.110**	0.070
	(0.032)	(0.055)
Deductible	-0.446	-0.152
	(0.290)	(0.231)
R-squared	0.203	
Adj. R-squared	0.199	
Num. obs.	20,340	

Note: * $p < 0.05$, ** $p < 0.001$. This table presents the results of a linear probability model that includes the interaction of each attribute with the teacher-quality index, the coefficients of which are presented in column (2).

TABLE 4—MITIGATING FACTOR IN DIFFICULT WORK SETTINGS

Outcome	LPM (1)	LPM (2)
Salary	0.061**	0.064**
	(0.002)	(0.002)
High Poverty	-0.067**	-0.107**
	(0.012)	(0.019)
Minority Share	0.002	0.002
	(0.002)	(0.002)
Low Achieving	-0.234**	-0.112**
	(0.021)	(0.009)
Principal Supportive	0.427**	0.526**
	(0.023)	(0.021)
Low Achieve × Support	0.253**	.
	(0.039)	.
High Poverty × Support	.	0.063*
	.	(0.030)
School bus blue	0.034**	0.027*
	(0.009)	(0.009)
Observations	*19,404*	*19,404*
R-squared	*0.366*	*0.363*

Note: * $p < 0.05$, ** $p < 0.001$. This table presents the results of two linear probability models in which I test whether having a principal "supportive with disruptive students" attenuates a teachers preference for richer and higher achieving school settings.

ONLINE APPENDIX A: 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 is $1.015. The cost of salary provision also interacts with the cost of salary growth and retirement, discussed below.

Growth

The choice experiment allows teachers to choose between varying salary growth rates. What makes the calculation of the cost of growth rates somewhat complicated is that providing more generous growth can reduce attrition, increasing the cost both through salaries at higher levels of experience *and* by increasing the odds that teachers make it to higher levels of experience. 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 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, for the labor market in question. I construct the average compensation cost in steady state, a function of compensation growth and increasing retention, as the salary growth number increases. Having done so for different values of g, I use a quadratic spline to generalize the cost of a one-unit increase in salary growth as:

$$\mathrm{cost_g} = (0.0707 \times \mathrm{salary} \times \mathrm{growth}) + (0.00000557 \times \mathrm{salary} \times \mathrm{growth}^2)$$

A thoughtful reader might wonder why we needn't account for how the attrition probabilities affect the cost of providing higher starting salary. The reason is that, at zero growth, the attrition pattern doesn't matter: The cost of a teacher's salary is always the same.

Health Insurance

In the July after this survey was administered (in May), I collected data from the 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 mostly affected 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.[31]

$$\mathrm{cost_d} = 12 \times (385.7 - 24.4 \times \mathrm{deductible})$$

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 will shift 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 compensation goes only to a quarter of teachers in the school, the cost of $1 merit pay is $0.25 per teacher. This income is subject to Medicare taxes, 1.45 percent.

[31] When the quadratic term is included, the coefficient's p-value is 0.688.

Defined Benefits Plan (Pension)

The calculation of pension costs is somewhat complicated. The explicit promise of a *defined benefits* program is that it is not subject to risk—the benefit, rather than just the contribution, is *defined.* Marx and Rauh (2014) show that, in order to satisfy the funding requirements, pension funds assume a constant high rate of growth (7.5–8.0 percent) to balance their revenues with their expected demands. This leads to underfunding above and beyond the underfunding recognized even under these rosy assumptions. The actual return of a risk-free investment, like treasury's, is 1.7 percent. I assume a rate of 2 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 online annuity calculator.[32] I then take the annual annuity as a fraction of the teacher's highest salary to make a mapping from what percent of salary the teacher is saving to her replacement rate. With a 2 percent risk-free rate of return, a one-percent saving pattern replaces two percent of the teacher's salary, meaning that teachers must save 0.510 to finance an additional percentage point of replacement rate under a risk-free rate of return.

Defined Contributions Plan (403(b))

Nonprofit and governmental agencies can provide a service corollary to the 401(k), called the 403(b), which are only available to tax-exempt organizations. In these 403(b)s, the school commits to contributing a defined amount to the worker's retirement rather than promising a defined level of benefit at retirement. While pensions take several years for a worker to vest,[33] 403(b) plans vest immediately, making the retirement contribution totally portable. I follow the same calculation to generate the cost of an average replacement rate through the 403(b), but use as the expected interest rate five percent, under the historical trend (eight percent) to be conservative and reflect the

[32] http://money.cnn.com/tools/annuities/

[33] 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.

expectations by many economists that growth in the future will be lower than that enjoyed as a result of previous technology improvements (Cowen 2011; Gordon 2016). Here from, the cost of saving enough to replace one percent of your salary on average is 0.324 percent of your salary. The cost of more generous replacement rates therefore interacts with salary generosity:

$$cost_r = sal(1 + .00324rep) + sal(.0707grt + .00000557grt^2) * (1 + .00324rep)$$

If one assumed an eight-percent return, the coefficient on rep would be .00197 rather than .00324.

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 structure 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 here 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 class size reductions is far from linear. The cost of additional class-size reductions increase quadratically as class size falls. To accommodate this tradeoff in the optimization sections, I conceptualize the cost function as a compensation structure nested in a class-size choice, allowing the cost structure of teacher pay to flexibly affect the cost of class size adjustments.

The other element affecting the cost of reduced class size is the fixed costs of employing an additional teacher, the primary cost of which is actually hiring and training. Barnes, Crowe, and Shaefer (2007) and Watlington et al. (2010) study the costs of turnover in terms of recruiting, screening, and training. The authors do an in-depth accounting exercise with five school districts and find that every new hire costs, on average, $11,891 dollars across the five districts. Because average teacher turns over every 6.13 years (the average years of experience in Hendricks (2014) and, coincidentally, the average years of experience in Aldine ISD), the yearly cost of hiring is $1,938 per teacher each year. I developed the model to reflect that the cost of additional teacher slots would decrease if retention is improved (Hendricks 2014). Simulating retention patterns for a one-percent increase in salary, I find that it would change the turnover dynamics to increase the average teacher tenure to 6.14 years. Thus the per-person costs are a function $11,891/(6.1+.00424(sal-50k)/.5k). Therefore, in practice, retention patterns don't change significantly enough to meaningfully alter the per-employee cost of hiring. I calculate other fixed costs of employment, but they are more 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 from the district). 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 B: ACHIEVEMENT PRODUCTION FUNCTION

First, I sought to understand what a teacher's welfare-maximizing compensation structure would look like, assuming her utility reflected her welfare. A second, important consideration in "optimal compensation" is what structure of pay maximizes student achievement rather than teacher satisfaction. 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.

$$P = h(t, r, k)$$

Where the parent's impact is a function of the time a parent allots to the 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 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 will be a linearization.

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 by taught by a novice who are 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 in 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, or declines in quality among veteran teachers, which would contradict the estimates used to train the VAM profile in experience. The value-add 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 12.

Class Size

Though the causal evidence on class size in the US is somewhat mixed, the typical read is that large class sizes reduce student student achievement (Jepsen and Rivkin 2009; Chingos 2013; Schanzenbach 2014). One of the most careful and convincing studies, Hoxby (2000), exploits natural variation arising from cohort sizes and class-size rules and finds no impact of class size on student achievement; a useful feature of her analysis is that she tests the effects of class size this school year on the impacts of test scores after summer break. Her null result compared to the positive results of other studies may capture the rapidity of fadeout for class-size induced achievement affects. 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. In contrast, 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. The experimental setting may alter teachers' incentives, since the results of a known experiment may influence future working conditions. Cho, Glewwe, and Whitler (2012) follow Hoxby while using new data and find that a ten-student reduction in class size by 0.04–0.05σ for students in elementary school, essentially identical to Kruger (1999). The evidence available tends to suggest class size does not matter 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-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.

Merit Pay

The evidence on merit 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). What is particularly striking is that the two prominent RCTs on merit pay (Springer et al. 2010 and Fryer 2013) generate null effects while each of those estimated in natural settings suggest positive effects, suggesting either bias in natural experiments or Hawthorne effects. The settings of each study differ enough to make comparison difficult. In many, schools implemented the reform with other supports, in others the incentives apply to school-wide or district-wide goals. Because of the programs similarity to the one posed to teachers in my survey and the setting is geographically comparable (from Texas), I use Imberman and Loveheim (2015) for a parameter value. They use the fact that grade-level incentives are stronger for smaller grades.

ONLINE APPENDIX C: 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:

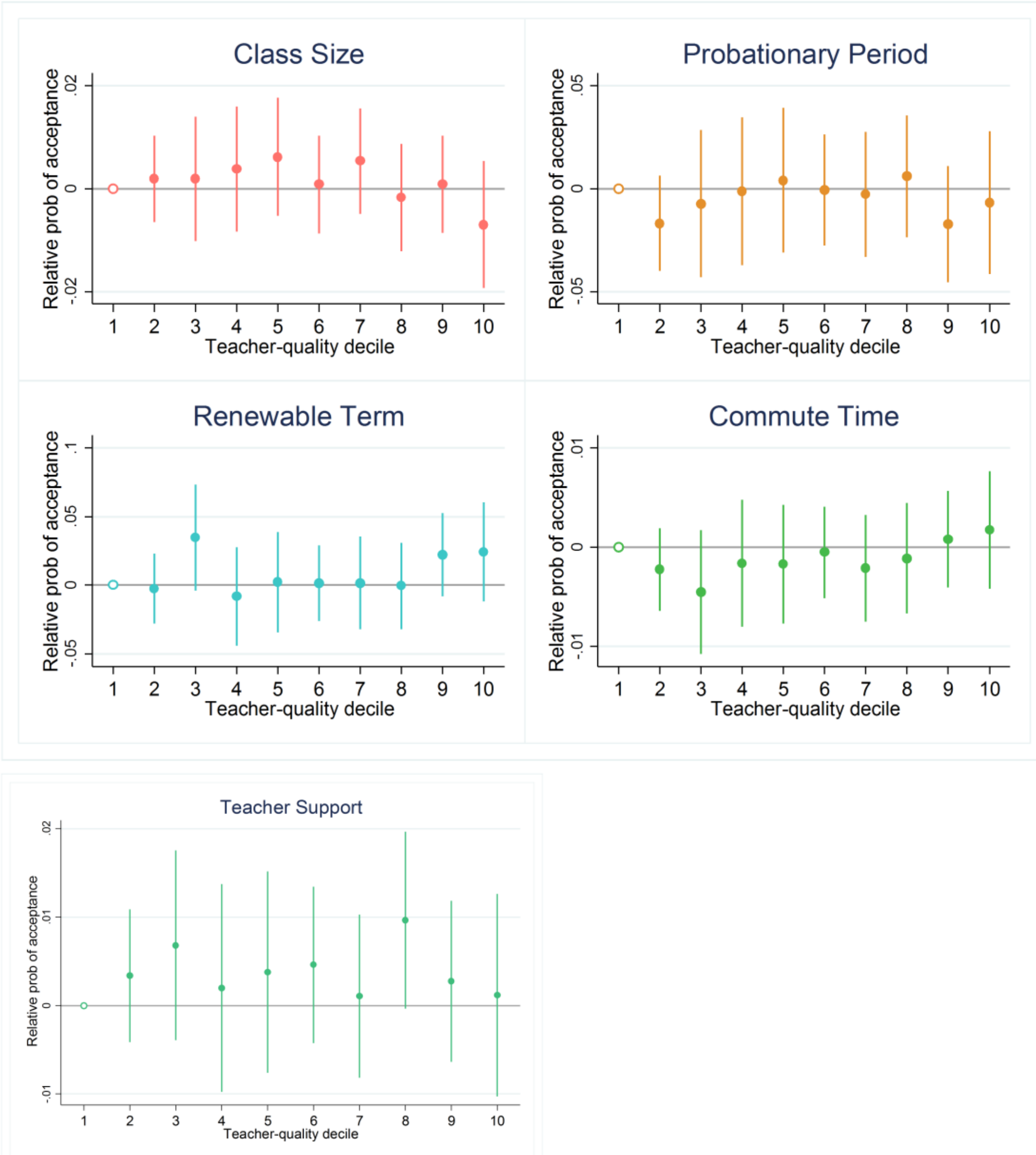
	School 1	School 2
Starting salary:	$47,150	$50,300
Percent of students in poverty:	38%	53%
Percent of students who are minority:	36%	66%
Average student achievement:	43rd percentile	57th percentile
Principal support:	Principals are hands-off with disruptive students	Principals are hands-off with disruptive 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.

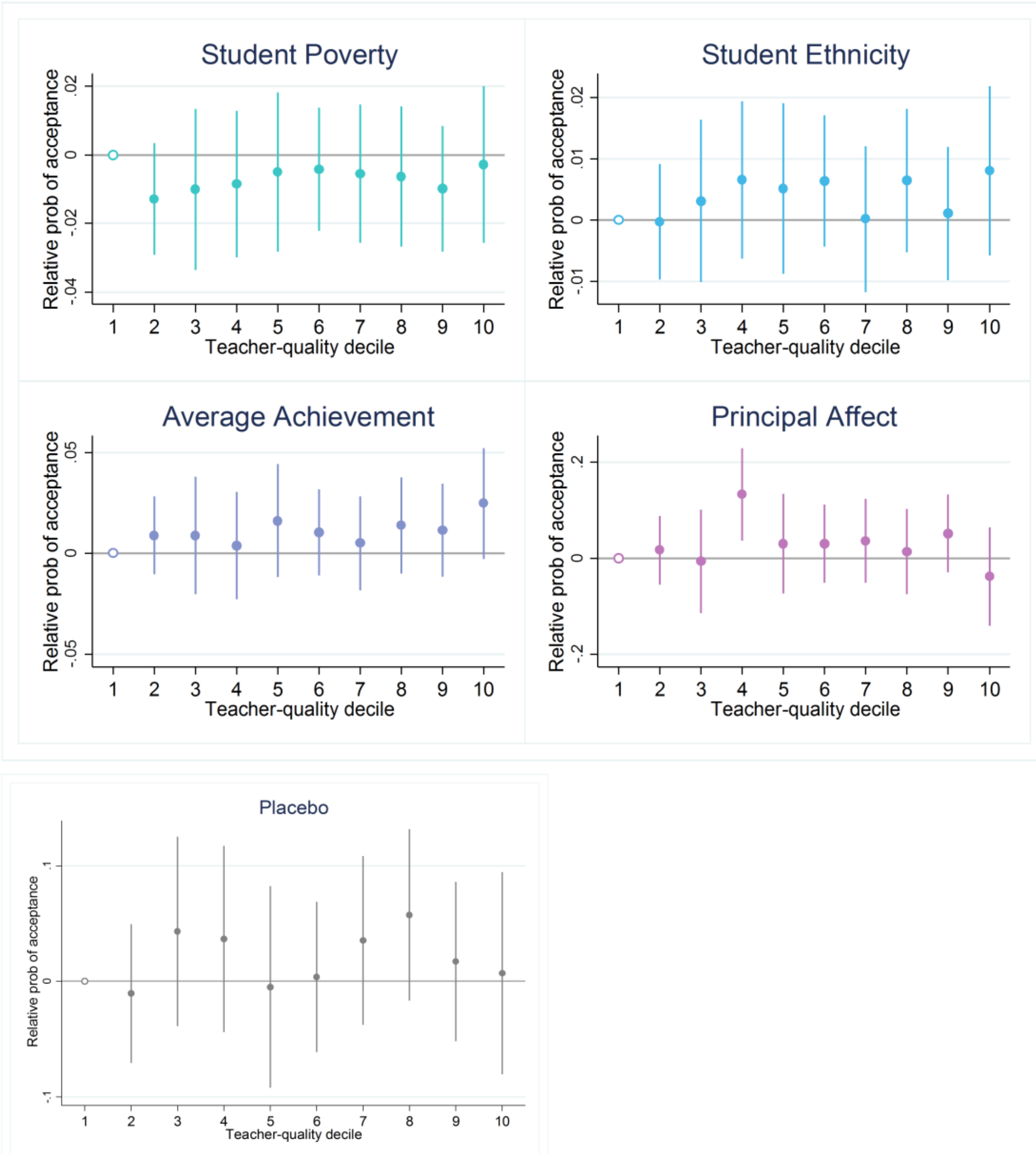
ONLINE APPENDIX FIGURE 5—DIFFERENTIAL WORKING-CONDITION PREFERENCE BY TEACHER-QUALITY DECILE



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

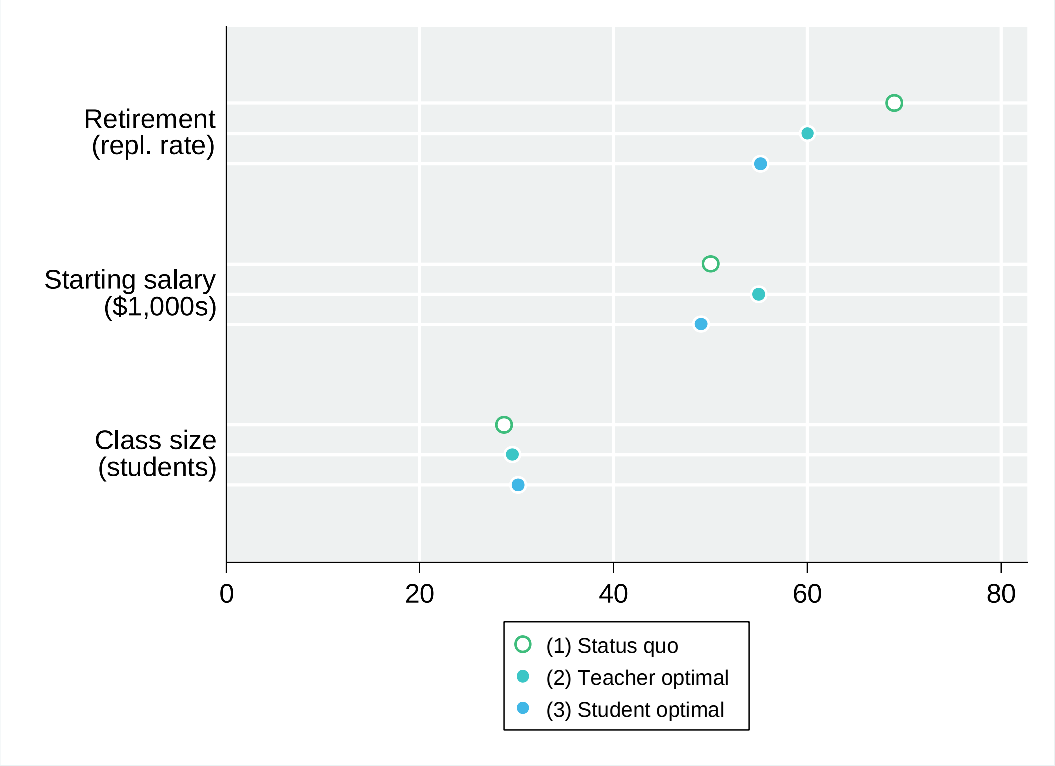
ONLINE APPENDIX FIGURE 6—DIFFERENTIAL STUDENTS-&-LEADERSHIP PREFERENCE

BY TEACHER-QUALITY DECILE



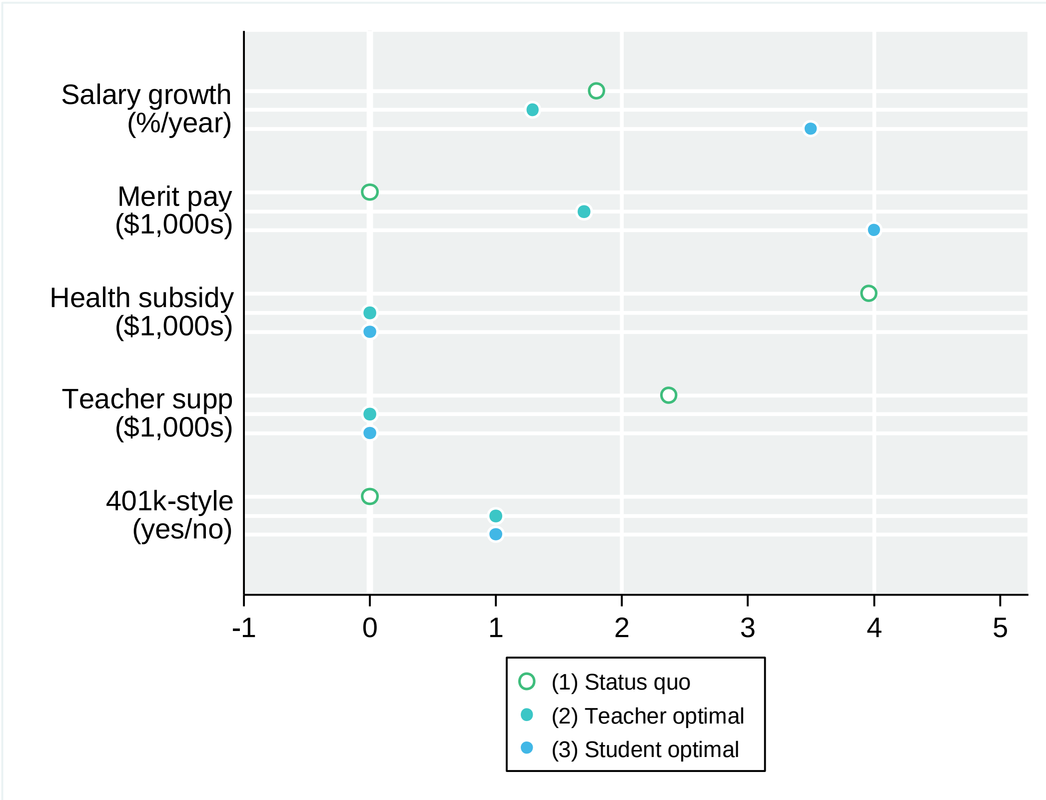
Note: This figure shows visually whether teachers of different quality deciles have distinct preferences for various student-and-leadership attributes, relative to bottom-decile teachers.

ONLINE APPENDIX FIGURE 7—COMPARING OPTIMAL CALCULATIONS OF COMPENSATION STRUCTURE AND MANIPULABLE WORKING CONDITIONS I



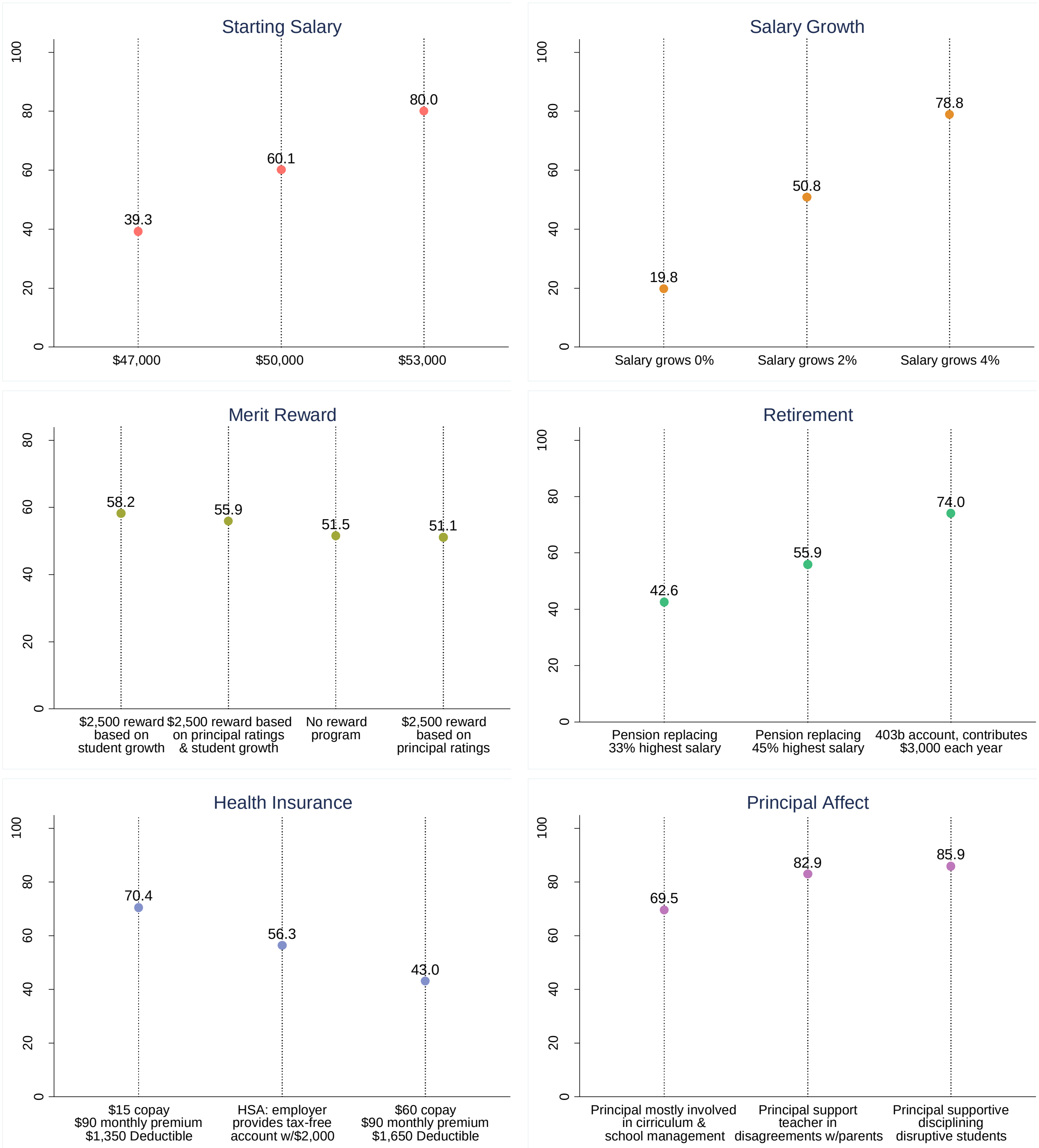
Note: This figure compares the status-quo compensation structure with teacher-utility maximizing compensation structure and student-achievement maximizing compensation structure.

ONLINE APPENDIX FIGURE 8—COMPARING OPTIMAL CALCULATIONS OF COMPENSATION STRUCTURE AND MANIPULABLE WORKING CONDITIONS II



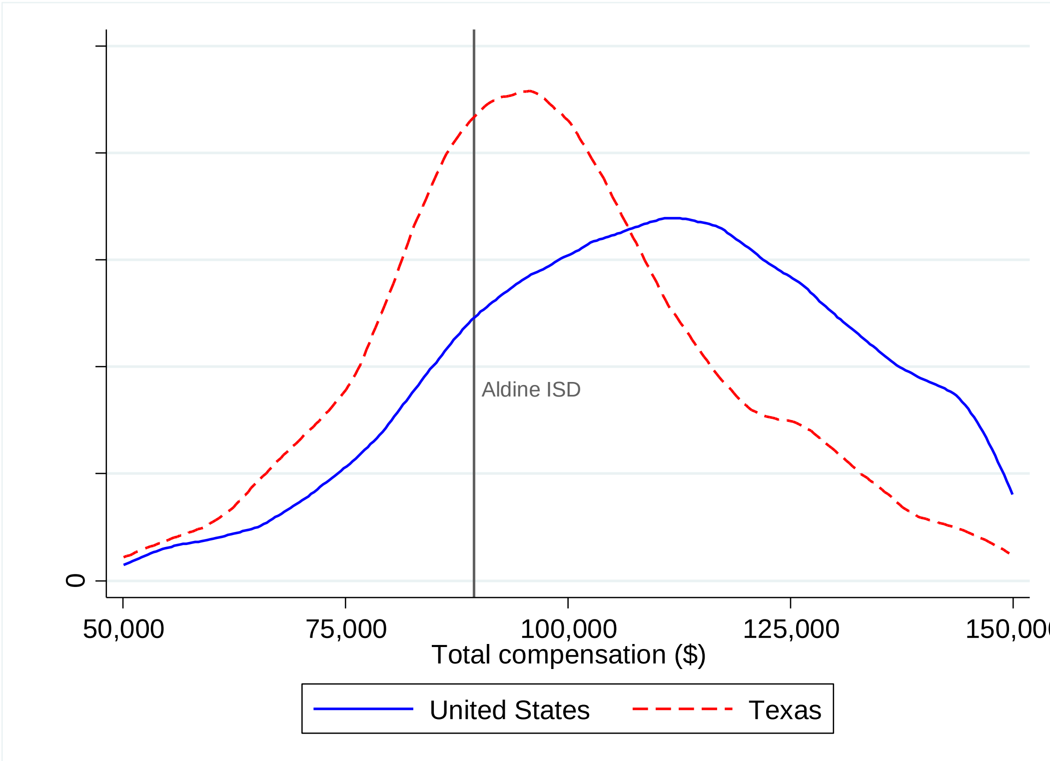
Note: This figure compares the status-quo compensation structure with teacher-utility maximizing compensation structure and student-achievement maximizing compensation structure.

ONLINE APPENDIX FIGURE 9—STAND-ALONE ATTRIBUTE EVALUATION QUESTION



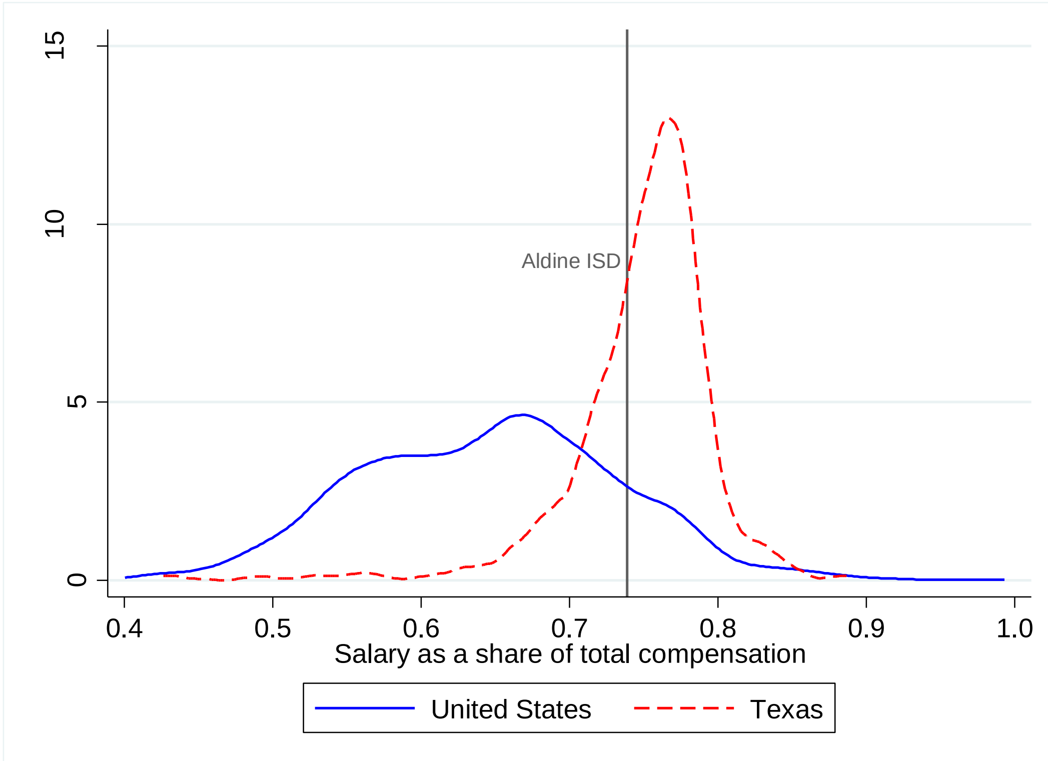
Note: This figure presents the results of additional survey questions in which a subset of teachers were asked to evaluate the probability that they would accept an offer that featured varying attributes.

ONLINE APPENDIX FIGURE 10—COMPARING ALDINE-ISD TOTAL COMPENSATION TO DISTRIBUTION



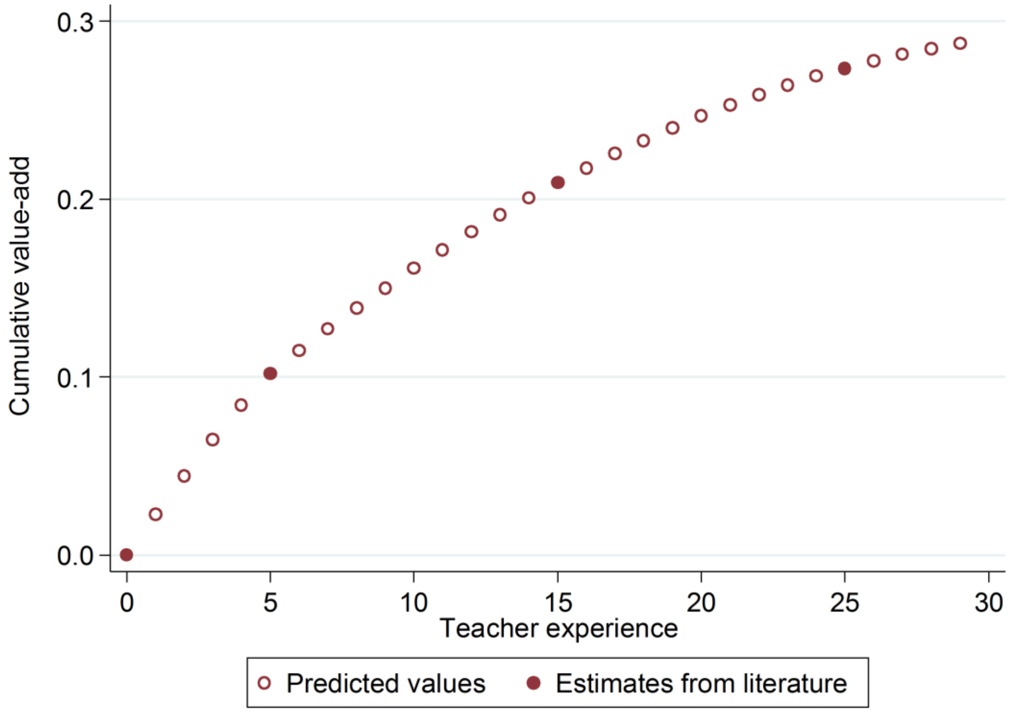
Note: This figure compares the average total compensation at Aldine ISD to the distribution of total compensation in the U.S. and in Texas using data from the Local Education Finance Survey.

ONLINE APPENDIX FIGURE 11—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 12—VALUE-ADD GROWTH WITH EXPERIENCE



Note: This figure shows the value-add 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 D: 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— 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	0.261***	(0.013)	0.000
$54,000	0.462***	(0.017)	0.000
Salary growth			
1 percent	0.193***	(0.017)	0.000
2 percent	0.378***	(0.019)	0.000
Merit pay			
$2000	0.114***	(0.015)	0.000
$3000	0.107***	(0.013)	0.000
VAM only	0.028	(0.019)	0.139
Retirement			
Replaces 40%	0.118***	(0.025)	0.000
Replaces 50%	0.265***	(0.035)	0.000
Replaces 60%	0.430***	(0.026)	0.000
Replaces 70%	0.503***	(0.032)	0.000
401k-style	0.146***	(0.019)	0.000
Health insurance			
$50/mo. premium	0.058***	(0.010)	0.000
$1,300 deductible	0.163***	(0.015)	0.000
R-squared	0.1971		
Adj. R-squared	0.1957		
Num. obs.	26,074		

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 3— 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	-0.212***	(0.036)	0.000
28 students	-0.458***	(0.030)	0.000
Probationary period			
1-year	-0.114***	(0.021)	0.000
2-year	-0.096***	(0.020)	0.000
3-year	-0.208***	(0.021)	0.000
Renewable terms			
2-year	-0.063***	(0.018)	0.000
3-year	-0.071***	(0.017)	0.000
Commute time			
~10 minutes	-0.005	(0.013)	0.678
~20 minutes	-0.050***	(0.014)	0.000
Teacher support			
5 hours/wk	0.115***	(0.016)	0.000
7 hours/wk	0.173***	(0.017)	0.000
9 hours/wk	0.241***	(0.018)	0.000
R-squared	0.2912		
Adj. R-squared	0.2899		
Num. obs.	25,870		

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 4— 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	-0.073***	(0.023)	0.001
80% low-income	-0.087***	(0.019)	0.000
100% low-income	-0.158***	(0.027)	0.000
Student ethnicity			
40% minotiry	0.001	(0.036)	0.976
70% minotiry	0.001	(0.041)	0.982
100% minotiry	0.009	(0.033)	0.793
Average achievement			
50th percentile	0.155***	(0.015)	0.000
66th percentile	0.373***	(0.038)	0.000
Principal affect			
Supportive	0.769***	(0.014)	0.000
Placebo			
Bus blue	0.010	(0.012)	0.406
R-squared	0.3666		
Adj. R-squared	0.3652		
Num. obs.	19,404		

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 5—PREFERENCES FOR WORKING CONDITIONS BY TEACHER QUALITY

	Choice	
	Reference Group	Quality-decile interaction
	(1)	(2)
Benchmark		
Starting salary	0.118***	-0.002
	(0.002)	(0.002)
Contract		
Probationary period	-0.063***	0.012
	(0.008)	(0.012)
Term length	-0.013	0.021
	(0.009)	(0.013)
Working conditions		
Commute time	-0.007***	0.002
	(0.001)	(0.002)
Class size	-0.071***	0.000
	(0.003)	(0.004)
Assistance	0.027***	0.002
	(0.002)	(0.004)
R-squared	0.289	
Adj. R-squared	0.286	
Num. obs.	20,274	

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents the results of a linear probability model that includes the interaction of each attribute with the teacher-quality index, the coefficients of which are presented in column (2).

ONLINE APPENDIX TABLE 6—PREFERENCES FOR STUDENT AND LEADERSHIP CHARACTERISTICS BY TEACHER QUALITY

	Choice	
	Reference Group	Quality-decile interaction
	(1)	(2)
Benchmark		
Starting salary	0.068**	-0.003
	(0.003)	(0.002)
Students		
Percent low income	-0.026**	0.004
	(0.005)	(0.008)
Percent minority	0.000	0.005
	(0.003)	(0.005)
Ave. achievement	0.027**	0.011
	(0.006)	(0.009)
Principal affect		
Supportive	0.581**	0.007
	(0.021)	(0.035)
Placebo		
Blue bus	-0.015	0.035
	(0.017)	(0.029)
R-squared	0.372	
Adj. R-squared	0.368	
Num. obs.	15,202	

Note: * $p < 0.05$, ** $p < 0.001$. This table presents the results of a linear probability model that includes the interaction of each attribute with the teacher-quality index, the coefficients of which are presented in column (2).