ONLINE APPENDICES

Preferences, Selection, and the Structure of Teacher Pay

Andrew C. Johnston

Online Appendix A: Background on Conjoint Studies/Choice Experiments

The Empirical Challenge

When economists set out to estimate preferences, they collect data on the choices people make and the options available to them, sometimes called "menus." Unfortunately, the records needed to construct menus from which teachers select offers are unavailable. Districts have no reason to keep records of offers made, and—because of the structure of the market—teachers tend not to receive competing offers.[1] If these records were collected, omitted variables would make it impossible to isolate the causal effect of each attribute. As an example, salary would appear to be more preferred than it really is if schools that pay more also had better amenities. Alternatively, salary would appear less preferred than it really is if schools pay more to compensate for unobservably difficult work settings (Antos and Rosen 1975). In either case, the resulting estimates would not be useful for predicting the effect of policy experiments.

Even if these challenges were somehow surmountable, the results would not be particularly informative. There is essentially no independent variation in most of the school attributes that form a work setting for teachers. It is common for competing schools to have identical compensation structures, tenure timelines, and rules governing working conditions like class size. Even across districts, variation is extremely limited by market concentration, statewide policy, and the common effect of union bargaining (Biasi 2021). Districts within a state share a pension program. Where variation sometimes exists at the borders between districts, the wealthier district usually offers a work setting that exceeds the neighboring district in every dimension, providing no information on preferences other than what was already known: that more is usually preferred.[2]

How, then, can we study teacher preferences? I use a discrete-choice experiment in the field. I generate hypothetical job offers that randomly vary compensation structure and working conditions and measure teacher choice. An example of the questions asked is presented as figure 1. The experiment neatly addresses the empirical challenges endemic to the question. First, the setting allows us to directly observe menus so that we can see the options from which teachers

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

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

select. Second, the experiment addresses omitted variables using a controlled experimental setting in which there are no factors unobserved. And third, the environment allows me to introduce independent variation in important policy variables that don't exist or vary in the natural world. These are precisely the issues that make the study of teacher preferences challenging (and in some cases impossible) with naturally occurring records.

Evaluating the Validity of Discrete-Choice Experiments

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

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

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

Third, incentive compatibility seems to matter only when discovering one's preferences requires significant effort, or if subjects have a distinct reason to dissemble;[3] estimates from hypothetical choices align with those from incentivized elicitations in settings where respondents already know their preferences (Camerer and Hogarth 1999; Mas and Pallais 2017; Maestas et al. 2018). Because compensation and working conditions affect a teacher's daily life, they have likely considered their preferences, suggesting no need for new effort to discover them. Several papers document that experimental valuations approach a neo-classical ideal as subjects gain experience in the setting (List 2003, 2004a, 2004b).

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

Fifth, the method avoids the influence of social-desirability bias. There is a large literature documenting that respondents may alter their answers to present socially desirable responses (Atkin and Chaffee 1972; Campbell 1981; Cotter, Cohen, and Coulter 1982; Finkel, Guterbock, and Borg 1991; Fisher 1993; Krosnick 1999). Surveys where an interviewer is not present conduce truth-telling (Legget et al. 2003; List, Berrens, Bohara, and Kerkvliet 2004; Alpizar et al. 2008). The online survey avoids these issues by providing the subject essentially anonymous privacy. Moreover, the survey design allows the subject to be honest by shrouding sensitive preferences. Subjects are presented a "long list" of attributes, and so they have multiple plausible justifications for any choice made in the conjoint setting (Karlan and Zinman 2012; Hainmueller, Hopkins, and

[3] 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."

Yamamoto 2014). If a teacher selects an offer with fewer minority students, for instance, she can point to any of the other attributes of the option as her rationale. Respondents enjoy privacy even from the researcher. The analyst cannot infer the preferences of any individual because each respondent makes fewer choices than there are factors (Lowes et al. 2017). Teacher responses are kept confidential and have been reliably private in previous surveys implemented by the consulting group with whom I partnered; thus, teachers have no reason to believe their employer will ever be able to review their individual responses.

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

References

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

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

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

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

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

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

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

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

Finkel, Steven E., Thomas M. Guterbock, and Marian J. Borg. “Race-of-Interviewer Effects in a Pre-Election Poll Virginia 1989.” *Public Opinion Quarterly* 55, no. 3 (1991): 313-330.
Fisher, Robert J. “Social Desirability Bias and the Validity of Indirect Questioning.” *Journal of Consumer Research* 20, no. 2 (1993): 303–315.
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.
Herriges, Joseph, Catherine Kling, Chih-Chen Liu, and Justin Tobias. “What Are the Consequences of Consequentiality?” *Journal of Environmental Economics and Management* 59, no. 1 (2010): 67-81.
Karlan, Dean S., and Jonathan Zinman. “List Randomization for Sensitive Behavior: An Application for Measuring Use of Loan Proceeds.” *Journal of Development Economics,* 98, no. 1 (2012): 71-75.
Krosnick, Jon A. “Maximizing Questionnaire Quality.” *Measures of Political Attitudes* 2 (1999): 37-58.
Leggett, Christopher G., Naomi S. Kleckner, Kevin J. Boyle, John W. Dufield, and Robert Cameron Mitchell. “Social Desirability Bias in Contingent Valuation Surveys Administered Through In-Person Interviews.” *Land Economics* 79, no. 4 (2003): 561-575.
List, John A. “Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards.” *American Economic Review* 91, no. 5 (2001): 1498-1507.
List, John A. “Does Market Experience Eliminate Market Anomalies?” *The Quarterly Journal of Economics* 118, no. 1 (2003): 41-71.
List, John A. “Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace.” *Econometrica* 72, no. 2 (2004): 615-625.
List, John A., Robert P. Berrens, Alok K. Bohara, and Joe Kerkvliet. “Examining the Role of Social Isolation on Stated Preferences.” *American Economic Review* 94, no. 3 (2004): 741-752.
Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan L. Weigel. “The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom.” *Econometrica* 85, no. 4 (2017): 1065-1091.
Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. "Truth in consequentiality: theory and field evidence on discrete choice experiments." *American Economic Journal: Microeconomics* 4, no. 4 (2012): 145-71.

Online Appendix B: Value-Added Measures

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

I estimate value added with an empirical Bayes shrinkage estimator following Kane and Staiger (2008). The process is outlined as follows:

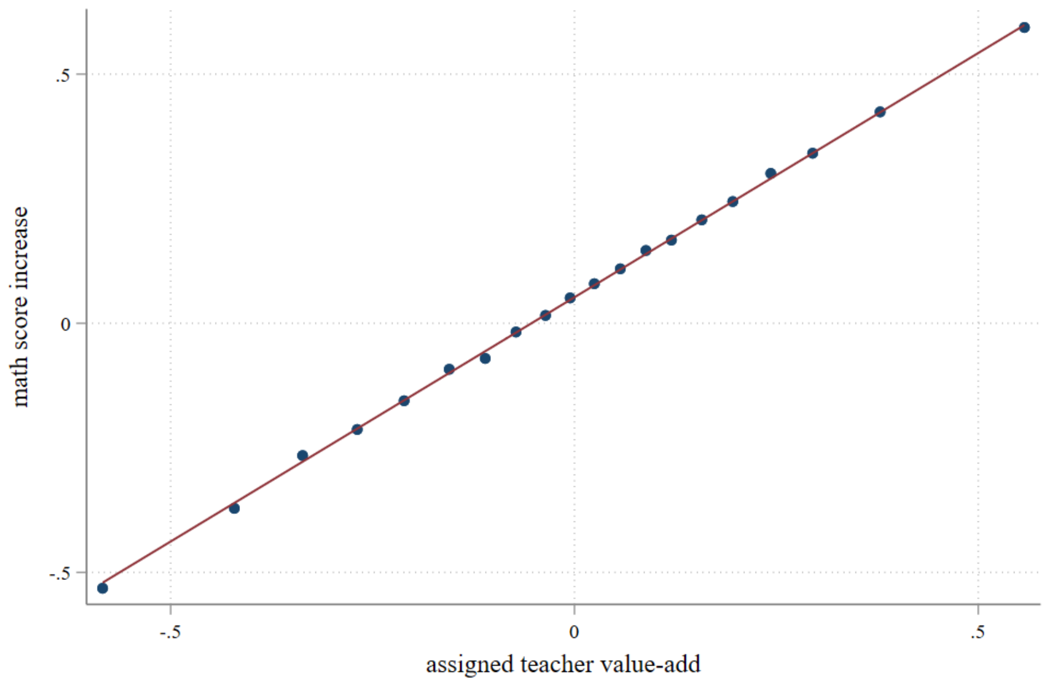
1. I standardize math and reading scores to have a mean of zero and a standard deviation of one.

2. I then estimate models of student achievement based on lagged achievement and other controls, and use those models to predict the residuals for each model in math and reading.
3. Then I calculate the mean residual of each teacher in each year (I do not see classroom).
4. Next, I calculate sigma-epsilon (the variance in the residuals not explained by teachers/classrooms) and sigma-mu (the covariance between teacher residuals in adjacent years).
5. Following, I calculate mean mean-residuals for teachers across years where the yearly mean residuals are weighted based on sigma-epsilon and the sample size of student test scores the teacher had in that year.
6. Finally, I "shrink" the weighted mean mean-residuals using sigma-mu, sigma-epsilon, and the sample size, essentially attenuating noisier estimates toward zero.

I estimate VA with six different sets of controls to scan for potential failures of robustness. The six models are:

- Model 1: linear pre scores in the same subject (math or reading); missing FE interacted with year; and school-year FE
- Model 2: linear and quadratic prescores in the same subject (math or reading); missing FE interacted with year; and school-year FE
- Model 3: linear, quadratic, and cubic prescores in the same subject (math or reading); missing FE interacted with year; and school-year FE
- Model 4: linear, quadratic, and cubic prescores in both subjects (math and reading); missing FE interacted with year; and school-year FE
- Model 5: linear, quadratic, and cubic prescores in the same subject (math or reading); year FE; missing FE interacted with year
- Model 6: linear, quadratic, and cubic prescores in both subjects (math and reading); year FE; missing FE interacted with year

The resulting VA estimates meet various reliability checks. I use Model 4 in the empirical analysis because it has the best fit and low prediction bias. The relationship between estimated VA and resulting test score increases are strong with a slope of approximately 1 (0.993 (0.0039) in mathematics and 1.009 (0.0062) in reading). Moreover, the VA measures add 7 percentage points to the R-Squared in a regression explaining student achievement. A 1SD increase in math VA is associated with a 0.16 SD increase in student math test scores. A 1 SD increase in reading VA is associated with a 0.11 SD increase in student reading test scores.



Notes: In this figure, I provide a binscatter plot of year-t test scores over assigned teacher value-add in year t*, while controlling for last year's test scores.*

The six models are similar in terms of model fit and prediction. Each model reproduces similar patterns with strong differential preference for performance pay as value add increases.

Online Appendix C: 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 section, I provide detail on how the cost function is constructed.

Salary

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

Health Insurance

In July 2016, three months after the survey was administered, I collected data from the Affordable Care Act (ACA) health exchange which indicated the monthly premium, deductible, cost of an office visit, and plan type (HMO, PPO, POS, PD, catastrophic) for 50 plans available in the Houston area. A hedonic pricing model revealed that the cost of office visits (the copay) had no systematic relationship with price (premia), which was most predicted by the deductible ($p < 0.001$) and HMO status ($p < 0.001$). With no deductible, a generic plan cost \$385.70 (CI: \$361.34 – \$410.06) per month, and the cost declined by \$24.40 (\$20.30 – \$28.49) for every \$1,000 increase in the deductible. There is no evidence that the price is a quadratic function of the deductible.[4]

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

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

Merit Pay

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

Defined Benefits Plan (Pension)

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

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

accumulated by retirement (assumed at age 65) and annuitize it, using an annuity calculator.[5] 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 1.57 percent risk-free rate of return, a one-percent saving pattern replaces two percent of the teacher's salary, meaning that teachers must save 0.559 percent of their income to finance an additional percentage point of replacement rate under a risk-free rate of return. Pensions, however, enjoy a cost saving since some teachers will pay into the pension but will not persist long enough to vest and receive an annuity. I calculate the share of those paying into the pension each year who will leave *before* the vesting period is complete. That fraction is then applied as a discount on the cost of the pension.

Defined Contributions Plan (403(b))

Nonprofit and governmental agencies can provide a retirement plan to their employees similar to the 401(k), called the 403(b), which are available to all tax-exempt organizations. In 403(b) accounts, the school commits to contributing a defined amount to the worker's retirement rather than promising a defined level of benefits at retirement. While pensions take several years for a worker to vest and retirement benefits are heavily backloaded,[6] 403(b) plans accumulate retirement wealth proportional to employment and vest immediately, making retirement contribution totally portable. I follow the same calculation as described above to generate the cost of an average replacement rate through the 403(b), and use as the expected interest rate 7.5 percent, under the historical trend (ten percent) (Cowen 2011; Gordon 2016). The cost of saving to replace one percent of a teacher's salary in expectation is 0.220 percent of your salary. If one assumed an eight-percent return, the coefficient on rep would be 0.202 percent.

Class Size

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

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

[6] 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.

structure (which determines the average cost per teacher) and class size (which determines the number of teachers needed), allowing the cost structure of teacher pay to flexibly affect the cost of class-size adjustments. To provide a smooth function for optimizing, I model teacher quantity as continuous.

Endogenous Retention

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

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

Cost of Turnover

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

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

Online Appendix D: Objective Functions

Teacher Utility

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

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

Teacher Retention

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

Student Achievement Production Function

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

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

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

Experience

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

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

To provide a general profile of experience on quality, I average the math and English returns. I fit a regression model of average VAM on experience and experience-squared using the first three experience nodes (0, 5, and 15), and a second model using the latter three points (5, 15, and 25) and use the predicted values (y-hat) from 0 through 5 in the first model and between 6 and 30 in the second model. Without the combination of these two piecewise models, the resulting experience profile either suggests convex increases in quality among veteran teachers—something never found in empirical work—or declines in quality among veteran teachers, which would contradict the estimates used to train the VAM profile in experience. The value-added profile that results from this procedure is most steeply increasing for new teachers but reflects the gains of experience throughout the life cycle of a teacher (Wiswall 2013; Papay and Kraft 2015). The resulting performance profile is presented in online Appendix figure 4.

Class Size

Analysts typically conclude that large class sizes reduce student achievement, especially for students that are young or low-income (Angrist and Lavy 1999; Krueger and Whitmore 2001; Jepsen and Rivkin 2009; Fredriksson, Ockert, Oosterbeek 2012, 2016; Schanzenbach 2014), but the literature also contains a significant corpus finding precise null effects (Hoxby 2000; Chingos 2013; Angrist, Lavy, Leder-Luis, and Shany 2019). In this paper, I incorporate domestic estimates of the influence of class size into the education production function. Krueger (1999) finds that an eight-student reduction (from 23 students to 15) increased achievement by 0.035σ per year, with larger effects in kindergarten (0.120σ), using random assignment from the Tennessee STAR experiment.[7] In contrast, Hoxby (2000) exploits natural variation arising from cohort sizes and class-size rules and finds no impact of class size on student achievement; her use of test scores after summer break may reveal rapid fadeout for class-size induced achievement gains. Dee and West (2011) use a within-student comparison for middle-school students and, similarly, find no overall impact of class size on student achievement. Cho, Glewwe, and Whitler (2012) follow Hoxby using recent data and find that a ten-student reduction in class size increases achievement by 0.04–0.05σ for students in elementary school, essentially in line with Krueger (1999). The domestic evidence tends to suggest class size 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.

Performance Pay

The evidence on performance pay suggests modest improvements to achievement in the presence of stronger incentives, but this literature is also split (Lavy 2002; Springer et al. 2010; Muralidharan and Sundararaman 2011; Sojourner, Fryer et al. 2011; Fryer 2013; Mykerezi, and West 2014; Dee and Wyckoff 2015; Imberman and Lovenheim 2015; Balch and Springer 2015). The settings of each study differ enough to make comparison difficult. In many programs, schools implemented the reform with other supports; in others, the incentives apply to school-wide or district-wide goals. Because of the program's similarity to the one posed to teachers in my survey and the setting is geographically proximate (from Houston, Texas), I use Imberman and Loveheim

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

(2015) for a parameter value. They use the fact that grade-level incentives are stronger for smaller grades and find that a \$1,000 merit-pay increase induces a 0.0136σ increase in student achievement (approximately the average of the math and English effect using specifications 1 and 2 in table 4 for a 10 percentage point increase in exposure (effect divided by 10), divided by the typical payment in thousands, \$1.283). This is a conservative parameter value since the first 10-percent increment exposure has a much larger effect (between 0.05 and 0.09 student standard deviations).

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

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

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

Online Appendix E: The Effects of Compensation Reform in General Equilibrium

The baseline simulation is partial equilibrium: what would happen if a single district implemented the optimal reform. It is useful to consider how these effects would scale in general equilibrium—if all schools adopted the optimal reform. Some of the effects in the partial equilibrium calculation will apply directly in general equilibrium. For instance, the effect of class size on achievement exists no matter how many districts implement personnel adjustments. The idea is that the effect of class size is direct (not mediated by allocation) and one district implementing compensation/class-size changes does not affect the productivity of another district changing its class size. Implementing reform in one district has no effect on the impact of class size in another. The same logic applies to the effect of performance incentives on effort. These inputs do not affect achievement through the reallocation of scarce resources among districts, and therefore they have the same effect in partial and general equilibrium.

Partial and general equilibria depart one another in the domain of district retention and selective retention. (For simplicity, we have ignored the effects of selection on entry other than to show that compensation preferences are indistinguishable for groups more and less disposed to teaching, which implies the stability of the optima.) The key is to understand to what extent compensation-induced retention at a district retains teachers who would have otherwise gone to another district, and to what extent compensation-induced retention at the district retains those who would have otherwise left public school teaching in Texas.

I collect staffing data from Texas that cover all public-school employees in the state from 2005 to 2015. The data include the base pay, education, experience, district, and a unique teacher identification code for each staffing record. I impute when a teacher leaves a district when they have stopped working for a district for at least three years and begin working at a new one. I impute that a teacher has departed public-school teaching when they have stopped working in public schools in Texas for at least three years. I recover the salary schedule of each district in each year by calculating the modal base salary for each experience cell in every district among full-time teachers for whom we have a record of them having a bachelors degree but not a masters.

When a teacher disappears from a district two outcomes are possible. One, the teacher has kept teaching but moved to another district. Two, the teacher has retired from public-school teaching in Texas. The method I pursue is to estimate the effect of salary changes on district exit and professional exit. I follow Hendricks (2014) who implements a clever strategy exploiting changes in salary schedules that vary by district, experience level, and time. This permits a rich, saturated set of controls including year-district fixed-effects, year-experience fixed-effects, and district-experience fixed effects:

$$E_{idst} = \beta \times S_{dst} + \theta_{dt} + \alpha_{st} + \gamma_{ds} + \varepsilon_{idst}$$

Where *E* indicates the exit of a teacher *i* in district *d* at experience *s* at time *t*, and we measure two types of exit: that in which a teacher moves to another government school district in Texas, and that in which a teacher leaves government-school teaching completely. θ_{dt} denotes a set of district-year fixed-effects, α_{st} denotes a set of experience-year fixed-effects, and γ_{ds} denotes a set of district-experience fixed-effects. *S* is the salary paid to teachers in district *d* at experience-level *s* at time *t*, in $1,000s of dollars. Therefore the β captures the relationship between a $1,000 increase in salary on the probability of exit as a percentage point. Intuitively, the strategy leverages within-district comparisons where one experience rung has a raise relative to another experience rung in the same district at the same time. To gauge the plausibility of the estimates and estimate the dynamic effects of compensation on retention, the main specification I run is a distributed lag model in which I include four leads and four lags of *S* into the model as well as the contemporaneous effect. I cluster the standard errors by teacher. The results are presented in online Appendix table 29.

The contemporary effects of compensation are largest, and smaller effects exist immediately before the raise (anticipatory effects) and immediately after the raise (satisfaction effects). To produce a simple number reflecting the impact of compensation on the two types of retention, I sum the contemporary effect with leads and lag effects and compute the standard error of the composite using the delta method. A $1,000 increase in district salary reduces exit to other districts by 0.028 percentage points (t-statistic: 0.137), and reduces exit from the profession by

0.168 percentage points (on a base of 9.29 percent; *t*-statistic of 5.41). In other words, 13 percent (0.028 / (0.028 + 0.168) = 13 percent) of retentions induced by compensation changes in an individual district are the result of retaining teachers who would have transferred to another district, and 100 percent (0.168/ (0.028 + 0.168) = 87 percent) of retentions induced by compensation changes in an individual district are the result of retaining teachers who would have left the profession completely. In practice, none of the retentions induced by higher salaries come at the expense of other districts, since they only retain teachers who would have otherwise departed teaching. To calculate what portion of the partial equilibrium effect would be seen in general equilibrium, we sum the part coming from class size and the effort effects of incentives (88 percent) plus 87 percent of the effects from retention (12 percent) which yields 98.44 percent of the partial equilibrium gains would be seen in general equilibrium ($0.88 + 0.87 \times 0.12 = 98.44$). According to this analysis, all the partial equilibrium effects flow through in general equilibrium because induced retention is not at the expense of other districts.

Online Appendix F: 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:	2.0% each year	4.0% each year
Reward:	Teachers receive $1,000 reward if they are in the top 25% of the school based on principal ratings and student growth	Teachers receive $2,000 reward if they are in the top 25% of the school based on principal ratings and student growth
Retirement:	A pension that replaces 65% of your salary in retirement if you stay 30 years	A pension that replaces 35% of your salary in retirement if you stay 30 years

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

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.

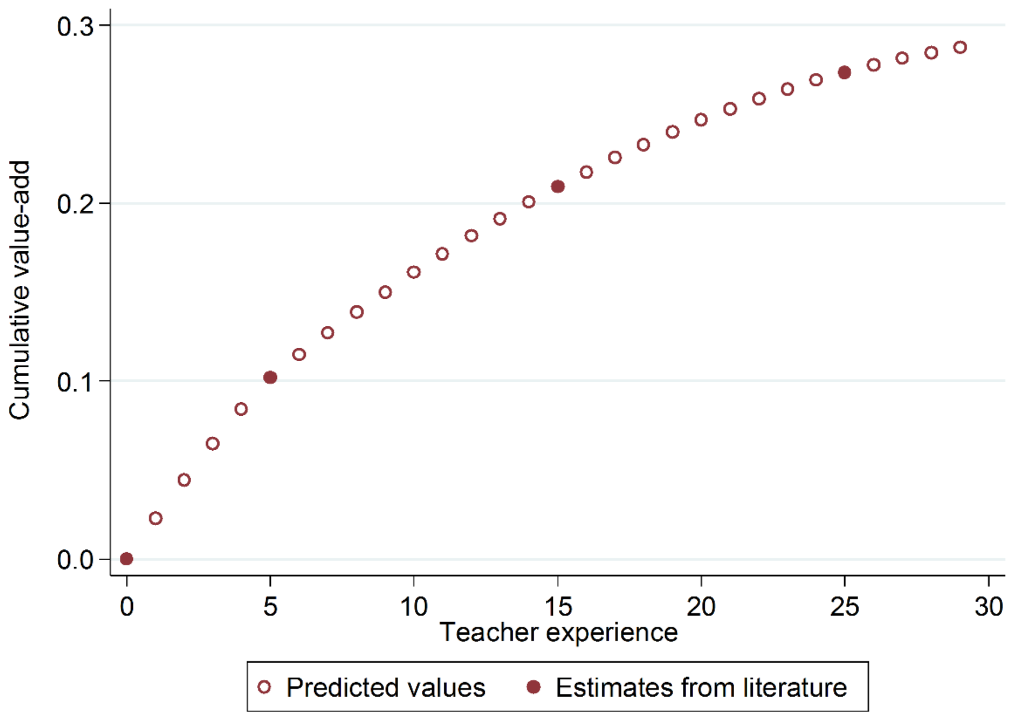
ONLINE APPENDIX FIGURE 3—SAMPLE STUDENTS-&-LEADERSHIP QUESTION

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
	○	○

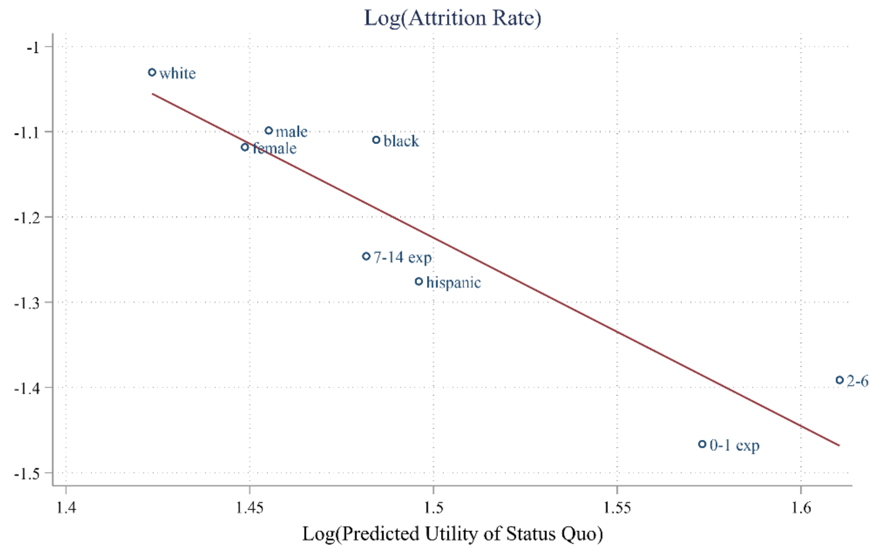
Note: This figure presents an illustration of the questions answered by teacher respondents with respect to student and principal characteristics.

ONLINE APPENDIX FIGURE 4—VALUE-ADDED GROWTH WITH EXPERIENCE



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

ONLINE APPENDIX FIGURE 5—GROUP ATTRITION AND GROUP UTILITY



Note: I calculate the average utility of the status quo bundle (in terms of salary, growth rates, replacement rates, deductibles, insurance subsidies, and class sizes) for eight groups: teachers with 0-1 years of experience, 2-6 years, 7-14 years, female teachers, male teachers, white teachers, black teachers, and Hispanic teachers. (I exclude the group of teachers with 15-36 years because these observations may be departing for retirement rather than low welfare.) I then calculate the relationship between their utility and their attrition rate, presented below. Consistent with a welfare improving retention, there is a strong negative relationship between a group's welfare from the current bundle and the group's attrition rate. For a 1 percent increase in utility, group attrition falls by 2.2 percent on average. The R-squared is 0.834 and the t-statistic on the slope parameter is 5.49. The model has nearly the same fit in levels.

ONLINE APPENDIX G: ONLINE APPENDIX TABLES

ONLINE APPENDIX TABLE 1—OFFER ATTRIBUTES FOR CONJOINT EXPERIMENTS

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

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

ONLINE APPENDIX TABLE 2 – TEACHER DEMOGRAPHICS

	Average	Std. Dev.
Experience in years	8.923	(9.160)
Bachelor's	0.440	(0.496)
Master's	0.285	(0.451)
White	0.266	(0.442)
Hispanic	0.200	(0.400)
Black	0.353	(0.478)
Female	0.655	(0.475)
Math VA	0.047	(0.305)
Reading VA	0.006	(0.184)
Danielson score	12.756	(2.137)

Note: This table presents the demographic makeup of teacher respondents in a dataset in which each row represents a unique teacher respondent.

ONLINE APPENDIX TABLE 3—PREFERENCES FOR WORKING CONDITIONS BY TEACHER QUALITY

	Choice (1)	Choice (2)	Choice (3)
salary	0.119	0.119	0.115
	(0.003)	(0.003)	(0.002)
salary x quality	0.012	0.007	0.000
	(0.007)	(0.011)	(0.001)
probationary period	-0.060	-0.059	-0.053
	(0.009)	(0.009)	(0.006)
prob. x quality	0.024	0.042	-0.001
	(0.017)	(0.027)	(0.002)
term length	0.005	0.005	-0.003
	(0.009)	(0.009)	(0.007)
term. x quality	-0.004	0.008	0.001
	(0.017)	(0.028)	(0.002)
commute time	-0.006	-0.006	-0.006
	(0.001)	(0.001)	(0.001)
commute. x quality	0.003	-0.000	0.000
	(0.003)	(0.005)	(0.000)
assistance	0.027	0.027	0.029
	(0.002)	(0.002)	(0.001)
assistance x quality	-0.002	-0.007	0.001
	(0.005)	(0.009)	(0.001)

Quality measure	Math VA	Read VA	Danielson
Question FE	X	X	X
R-squared	0.283	0.282	0.284
N	10,344	10,344	18,642

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

ONLINE APPENDIX TABLE 4—PREFERENCES FOR STUDENT AND LEADERSHIP CHARACTERISTICS BY TEACHER QUALITY

	Choice (1)	Choice (2)	Choice (3)
salary	0.066	0.067	0.063
	(0.003)	(0.003)	(0.002)
salary x quality	0.017	0.016	-0.000
	(0.010)	(0.017)	(0.001)
low-income %	-0.022	-0.022	-0.024
	(0.004)	(0.004)	(0.003)
low income x quality	-0.001	0.011	-0.001
	(0.011)	(0.017)	(0.001)
minority %	0.005	0.005	0.003
	(0.002)	(0.002)	(0.002)
minority x quality	0.001	-0.001	0.001
	(0.006)	(0.010)	(0.001)
student achievement	0.034	0.035	0.034
	(0.004)	(0.004)	(0.003)
achievement x quality	0.017	0.007	-0.000
	(0.012)	(0.022)	(0.001)
supportive principal	0.600	0.598	0.590
	(0.016)	(0.015)	(0.012)
supportive x quality	-0.055	-0.067	0.003
	(0.044)	(0.075)	(0.005)
blue bus	0.009	0.009	0.003
	(0.014)	(0.014)	(0.010)

blue bus x quality	0.020	0.093	0.001
	(0.040)	(0.070)	(0.004)
Quality measure	Math VA	Read VA	Danielson
Question FE	X	X	X
R-squared	0.383	0.383	0.377
N	7,726	7,727	13,970

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

ONLINE APPENDIX TABLE 5—LEAVER HETEROGENEITY IN COMPENSATION PREFERENCES

	Choice (1)	Choice (2)
Starting salary	0.085	0.085
	(0.002)	(0.002)
Starting salary x quit	-0.001	-0.002
	(0.001)	(0.002)
Salary growth	0.188	0.186
	(0.010)	(0.010)
Salary growth x quit	0.011	0.008
	(0.009)	(0.010)
Bonus amount	0.031	0.031
	(0.004)	(0.004)
Bonus amount x quit	-0.003	0.003
	(0.006)	(0.006)
VAM only	-0.069	-0.068
	(0.015)	(0.016)
VAM only x quit	-0.020	-0.017
	(0.013)	(0.015)
Replacement	0.014	0.014
	(0.001)	(0.001)
Replacement x quit	0.001	0.001
	(0.001)	(0.001)
401k-style	0.085	0.087
	(0.012)	(0.012)
401k-style x quit	-0.021	-0.023
	(0.015)	(0.016)

Experience FE	.	X
R-squared	0.193	0.196
N	31,820	28,456

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

ONLINE APPENDIX TABLE 6—LEAVER HETEROGENEITY IN WORKING CONDITION PREFERENCES

	Choice (1)	Choice (2)
Starting salary	0.116	0.118
	(0.002)	(0.002)
Starting salary x quit	-0.002	-0.002
	(0.001)	(0.001)
Probationary period	-0.063	-0.063
	(0.005)	(0.006)
Probationary x quit	0.013	0.011
	(0.006)	(0.007)
Term length	-0.002	-0.000
	(0.006)	(0.006)
Term length x quit	-0.005	-0.008
	(0.006)	(0.007)
Commute time	-0.006	-0.006
	(0.001)	(0.001)
Commute time x quit	-0.001	-0.001
	(0.001)	(0.001)
Assistance hours	0.028	0.028
	(0.001)	(0.001)
Assistance x quit	0.005	0.004
	(0.002)	(0.002)
Experience FE	.	X
R-squared	0.279	0.284

N	31,574	28,320

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

0

ONLINE APPENDIX TABLE 7—LEAVER HETEROGENEITY IN STUDENT AND PRINCIPAL PREFERENCES

	Choice (1)	Choice (2)
Starting salary	0.067	0.068
	(0.002)	(0.002)
Starting salary x quit	-0.001	-0.001
	(0.001)	(0.001)
Low income share	-0.022	-0.021
	(0.003)	(0.003)
Low income x quit	-0.000	-0.000
	(0.004)	(0.004)
Minority share	0.003	0.003
	(0.002)	(0.002)
Minority x quit	-0.001	0.001
	(0.002)	(0.003)
Achievement mean	0.033	0.033
	(0.003)	(0.003)
Achievement x quit	0.009	0.008
	(0.005)	(0.005)
Supportive principal	0.572	0.572
	(0.011)	(0.011)
Supportive x quit	0.009	0.010
	(0.017)	(0.018)
Blue bus	0.004	0.004
	(0.009)	(0.010)

Blue bus x quit	0.005	-0.005
	(0.014)	(0.016)
Experience FE	.	X
R-squared	0.365	0.368
N	23,678	21,246

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

ONLINE APPENDIX TABLE 8—COMPARING ESTIMATES FROM COLLEGE STUDENTS TO ALDINE ISD SAMPLE: COMPENSATION STRUCTURE

	Aldine ISD Teachers			U Houston Students			t-test of difference
	Coeff	SE	WTP	Coeff	SE	WTP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Salary							
Starting salary	0.085**	(0.002)	$1,000	0.043**	(0.004)	$1,000	9.39
Salary growth	0.192**	(0.009)	$2,268	0.275**	(0.012)	$6,319	-5.53
Merit reward							
Bonus amount	0.029**	(0.003)	$346	0.065**	(0.006)	$1,495	-5.37
VA only	-0.077**	(0.015)	-$907	-0.056*	(0.024)	-$1,293	-0.74
Retirement							
Replacement	0.015**	(0.001)	$173	0.014**	(0.001)	$316	0.71
401k-style	0.077**	(0.010)	$907	0.054**	(0.017)	$1,236	1.17
Health insurance							
Premium (yearly)	-0.082**	(0.014)	-$971	-0.072*	(0.024)	-$1,645	-0.36
Deductible	-0.312	(0.212)	-$3,686	-0.304**	(0.061)	-$6,989	-0.04
Observations	31,820			11,060			
R-Squared	0.193			0.174			

*Note: * $p < 0.05$, ** $p < 0.001$.* This table compares the preference estimates of the Aldine ISD sample to those of college students we survey nearby at the University of Houston. The models are estimated with least squares and the standard errors are clustered by respondent.

Online Appendix Table 9—Comparing Estimates from College Students to Aldine ISD Sample: Working Conditions

	Aldine ISD Teachers			U Houston Students			t-test of
	Coeff	SE	WTP	Coeff	SE	WTP	difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Salary							
Starting salary	0.115**	(0.002)	$1,000	0.084**	(0.004)	$1,000	6.93
Contract							
Probationary	-0.058**	(0.005)	-$501	-0.095**	(0.011)	-$1,130	3.06
Term length	-0.004	(0.005)	-$33	0.003	(0.011)	$37	-0.58
Working conditions							
Commute time	-0.365**	(0.043)	-$3,173	-0.781**	(0.095)	-$9,258	3.99
Class size	-0.068**	(0.001)	-$594	-0.031**	(0.003)	-$368	-11.70
Assistance	0.030**	(0.001)	$256	0.041**	(0.002)	$486	-4.92
Observations	31,574			8,346			
R-Squared	0.279			0.171			

*Note: * $p < 0.05$, ** $p < 0.001$.* This table compares the preference estimates of the Aldine ISD sample to those of college students we survey nearby at the University of Houston. The models are estimated with least squares and the standard errors are clustered by respondent.

ONLINE APPENDIX TABLE 10—COMPARING ESTIMATES FROM COLLEGE STUDENTS TO ALDINE ISD SAMPLE: STUDENT DEMOGRAPHICS AND PRINCIPAL SUPPORT

	Aldine ISD Teachers			U Houston Students			t-test of
	Coeff	SE	WTP	Coeff	SE	WTP	difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Salary							
Starting salary	0.066**	(0.002)	$1,000	0.033**	(0.004)	$1,000	7.38
Students							
% low income	-0.022**	(0.002)	-$325	-0.016**	(0.005)	-$494	-1.11
% minority	0.003	(0.001)	$40	0.007*	(0.003)	$221	-1.26
Ave. achievement	0.036**	(0.003)	$546	0.048**	(0.005)	$1,431	-2.06
Principal affect							
Supportive	0.575**	(0.009)	$8,672	0.332**	(0.017)	$10,000	12.63
Placebo							
Blue bus	0.007	(0.008)	$101	.	.	.	.
Dismissal							
Dismissal rate	.	.	.	0.185**	(0.027)	$5,570	.
Observations	23,678			8,490			
R-Squared	0.365			0.168			

*Note: * $p < 0.05$, ** $p < 0.001$.* This table compares the preference estimates of the Aldine ISD sample to those of college students we survey nearby at the University of Houston. The models are estimated with least squares and the standard errors are clustered by respondent. The survey of college students did not include the blue-bus question and instead varied the dismissal risk: essentially the percent of teachers that were dismissed for poor performance, where the values included 0.05%, 0.10%, 0.50%, or 1.00%.

ONLINE APPENDIX TABLE 11—DO COLLEGE-STUDENT PREFERENCES VARY WITH INTEREST IN TEACHING?: COMPENSATION STRUCTURE

	Choice		Choice	
	Coeff	SE	Coeff	SE
	(1)	(2)	(3)	(4)
Salary				
Starting salary	0.043**	(0.004)	0.042**	(0.004)
Starting salary x teaching index	.	.	0.001	(0.001)
Salary growth	0.275**	(0.012)	0.281**	(0.019)
Salary growth x teaching index	.	.	-0.003	(0.006)
Merit reward				
Bonus amount	0.065**	(0.006)	0.050**	(0.011)
Bonus amount x teaching index	.	.	0.006	(0.004)
VA only	-0.056*	(0.024)	-0.054	(0.032)
VA only x teaching index	.	.	0.000	(0.009)
Retirement				
Replacement	0.014**	(0.001)	0.014**	(0.002)
Replacement x teaching index	.	.	0.000	(0.000)
401k-style	0.054**	(0.017)	0.023	(0.029)
401k-style x teaching index	.	.	0.013	(0.010)
Health insurance				
Premium (yearly)	-0.072*	(0.024)	-0.022	(0.046)
Premium (yearly) x teaching index	.	.	-0.020	(0.016)
Deductible	-0.304**	(0.061)	-0.271*	(0.085)
Deductible x teaching index	.	.	-0.012	(0.023)
Observations	11,060		11,056	
R-Squared	0.174		0.174	

*Note: * $p < 0.05$, ** $p < 0.001$.* This table compares models of student preference that interact each attribute with a "teaching index" that measures interest-in-teaching. The teaching index varies from values of 1 ("I would never consider teaching") to 5 ("I plan to be a teacher"), where the average value is 2.5 (between "I've seriously considered teaching" and "I've thought about teaching") and the standard deviation is 1.3. The joint significance of the interaction terms has an F-statistic of 1.01 and a p-value of 0.419, suggesting preferences for teaching are not significantly related to preferences for compensation structure. The models are estimated with least squares and the standard errors are clustered by respondent.

ONLINE APPENDIX TABLE 12—DO COLLEGE-STUDENT PREFERENCES VARY WITH INTEREST IN TEACHING?: WORKING CONDITIONS

	Choice		Choice	
	Coeff	SE	Coeff	SE
	(1)	(2)	(3)	(4)
Salary				
Starting salary	0.084**	(0.004)	0.092**	(0.006)
Starting salary x teaching index	.	.	-0.003	(0.002)
Contract				
Probationary	-0.095**	(0.011)	-0.123**	(0.017)
Probationary x teaching index	.	.	0.011*	(0.005)
Term length	0.003	(0.011)	-0.014	(0.018)
Term length x teaching index	.	.	0.007	(0.006)
Working conditions				
Commute time	-0.781**	(0.095)	-0.819**	(0.164)
Commute time x teaching index	.	.	0.013	(0.056)
Class size	-0.031**	(0.003)	-0.022**	(0.005)
Class size x teaching index	.	.	-0.004*	(0.001)
Assistance	0.041**	(0.002)	0.039**	(0.004)
Assistance x teaching index	.	.	0.001	(0.002)
Observations	8,346		8,346	
Adjusted R-Squared	0.171		0.173	

Note: * $p < 0.05$, ** $p < 0.001$. This table compares models of student preference that interact each attribute with a "teaching index" that measures interest-in-teaching. The teaching index varies from values of 1 ("I would never consider teaching") to 5 ("I plan to be a teacher"), where the average value is 2.5 (between "I've seriously considered teaching" and "I've thought about teaching") and the standard deviation is 1.3. Students that are more interested in teaching have weaker aversion to probationary periods before tenure and stronger aversion to large class sizes. The joint significance of the interaction terms has an F-statistic of 2.51 and a p-value of 0.021. The models are estimated with least squares and the standard errors are clustered by respondent.

ONLINE APPENDIX TABLE 13—DO COLLEGE-STUDENT PREFERENCES VARY WITH INTEREST IN TEACHING?: STUDENT DEMOGRAPHICS AND PRINCIPAL SUPPORT

	Choice		Choice	
	Coeff	SE	Coeff	SE
	(1)	(2)	(3)	(4)
Salary				
Starting salary	0.033**	(0.004)	0.030**	(0.008)
Starting salary x teaching index	.	.	0.001	(0.003)
Students				
Percent low income	-0.016**	(0.005)	-0.031**	(0.009)
% low income x teaching index	.	.	0.006	(0.003)
Percent minority	0.007*	(0.003)	0.013*	(0.005)
% minority x teaching index	.	.	-0.002	(0.002)
Ave. achievement	0.048**	(0.005)	0.058**	(0.010)
Ave. achieve. x teaching index	.	.	-0.004	(0.003)
Principal affect				
Supportive	0.332**	(0.017)	0.249**	(0.033)
Supportive x teaching index	.	.	0.033*	(0.011)
Dismissal				
Dismissal rate	0.185**	(0.027)	0.211**	(0.042)
Dismissal rate x teaching index	.	.	-0.011	(0.013)
Observations	8,490		8,490	
R-Squared	0.168		0.171	

Note: $*p < 0.05$, $**p < 0.001$. This table compares models of student preference that interact each attribute with a "teaching index" that measures interest-in-teaching. The teaching index varies from values of 1 ("I would never consider teaching") to 5 ("I plan to be a teacher"), where the average value is 2.5 (between "I've seriously considered teaching" and "I've thought about teaching") and the standard deviation is 1.3. Students that are more interested in teaching have stronger preferences for supportive principals. The joint significance of the interaction terms has an F-statistic of 3.16 and a p-value of 0.004. The models are estimated with least squares and the standard errors are clustered by respondent.

ONLINE APPENDIX TABLE 14—EXPERIENCE HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability			
	Novice teachers (1st quartile: 0-1 yrs)	New-teacher differential (2nd quartile: 2-6 yrs)	Experienced-teacher differential (3rd quartile: 7-14 yrs)	Veteran-teacher differential (4th quartile: 15-36 yrs)
	(1)	(2)	(3)	(4)
Starting salary	0.098**	-0.002	-0.015*	-0.038**
	(0.003)	(0.005)	(0.005)	(0.005)
Salary growth	0.203**	-0.013	-0.023*	-0.024*
	(0.012)	(0.012)	(0.011)	(0.012)
Bonus amount	0.033**	0.002	0.006	-0.011
	(0.006)	(0.008)	(0.008)	(0.008)
VAM only	-0.068**	0.010	-0.005	-0.026
	(0.019)	(0.020)	(0.019)	(0.019)
Replacement	0.012**	0.002	0.003**	0.007**
	(0.001)	(0.001)	(0.001)	(0.001)
401k-style	0.087**	-0.018	0.005	-0.021
	(0.016)	(0.021)	(0.021)	(0.021)
Premium (yearly)	-0.073*	0.002	-0.003	-0.049
	(0.027)	(0.040)	(0.039)	(0.039)
Deductible	-0.466	0.161	0.180	0.474*
	(0.244)	(0.191)	(0.182)	(0.182)

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

ONLINE APPENDIX TABLE 15—EXPERIENCE HETEROGENEITY IN WORKING-CONDITION PREFERENCES

	Linear Probability			
	Novice teachers	New-teacher differential	Experienced-teacher differential	Veteran-teacher differential
	(1st quartile: 0-1 yrs)	(2nd quartile: 2-6 yrs)	(3rd quartile: 7-14 yrs)	(4th quartile: 15-36 yrs)
	(1)	(2)	(3)	(4)
Probationary period	-0.047**	-0.003	0.004	0.003
	(0.006)	(0.007)	(0.007)	(0.007)
Term length	-0.001	-0.010	0.000	0.002
	(0.006)	(0.007)	(0.007)	(0.007)
Commute time	-0.005**	0.001	0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Class size	-0.054**	0.000	-0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Assistance	0.019**	0.003	0.006*	0.007*
	(0.002)	(0.002)	(0.002)	(0.002)

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

ONLINE APPENDIX TABLE 16—EXPERIENCE HETEROGENEITY IN STUDENT/PRINCIPAL PREFERENCES

	Linear Probability			
	Novice teachers (1st quartile: 0-1 yrs)	New-teacher differential (2nd quartile: 2-6 yrs)	Experienced-teacher differential (3rd quartile: 7-14 yrs)	Veteran-teacher differential (4th quartile: 15-36 yrs)
	(1)	(2)	(3)	(4)
Percent low income	-0.030**	0.000	-0.001	0.008
	(0.006)	(0.008)	(0.007)	(0.007)
Percent minority	0.002	0.001	0.004	0.007
	(0.003)	(0.005)	(0.004)	(0.004)
Ave. achievement	0.043**	-0.000	-0.005	0.019*
	(0.006)	(0.009)	(0.009)	(0.009)
Supportive principal	0.707**	0.031	0.065*	0.138**
	(0.023)	(0.033)	(0.032)	(0.031)
Blue bus	-0.015	0.038	0.025	0.019
	(0.019)	(0.028)	(0.026)	(0.026)

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

Online Appendix Table 17—Sex Heterogeneity in Compensation Preferences

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

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

ONLINE APPENDIX TABLE 18—SEX HETEROGENEITY IN WORKING-CONDITION PREFERENCES

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

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

ONLINE APPENDIX TABLE 19—SEX HETEROGENEITY IN STUDENT AND PRINCIPAL PREFERENCES

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

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

ONLINE APPENDIX TABLE 20—RACIAL HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability		
	White teachers	Black differential	Hispanic differential
	(1)	(2)	(3)
Starting salary	0.081**	0.005	0.009*
	(0.003)	(0.004)	(0.004)
Salary growth	0.211**	-0.045**	-0.014
	(0.010)	(0.009)	(0.011)
Bonus amount	0.011*	0.037**	0.023*
	(0.005)	(0.006)	(0.007)
VAM only	-0.088**	0.030*	0.008
	(0.016)	(0.015)	(0.017)
Replacement	0.015**	-0.001	-0.002
	(0.001)	(0.001)	(0.001)
401k-style	0.059**	0.035*	0.024
	(0.013)	(0.016)	(0.019)
Premium (yearly)	-0.077**	-0.001	-0.019
	(0.021)	(0.030)	(0.036)
Deductible	-0.335	0.095	-0.074
	(0.222)	(0.144)	(0.167)

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

ONLINE APPENDIX TABLE 21—RACIAL HETEROGENEITY IN WORKING-CONDITION PREFERENCES

	Linear Probability		
	White teachers	Black differential	Hispanic differential
	(1)	(2)	(3)
Probationary period	-0.039**	-0.016*	-0.003
	(0.005)	(0.005)	(0.006)
Term length	0.000	-0.009	-0.000
	(0.005)	(0.006)	(0.007)
Commute time	-0.005**	0.001	0.001
	(0.001)	(0.001)	(0.001)
Class size	-0.056**	0.008**	-0.005*
	(0.001)	(0.002)	(0.002)
Assistance	0.023**	0.001	-0.001
	(0.001)	(0.002)	(0.002)

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

ONLINE APPENDIX TABLE 22—RACIAL HETEROGENEITY IN STUDENT AND PRINCIPAL PREFERENCES

	Linear Probability		
	White teachers	Black differential	Hispanic differential
	(1)	(2)	(3)
Percent low income	-0.032**	0.009	-0.001
	(0.004)	(0.006)	(0.007)
Percent minority	-0.000	0.011*	-0.001
	(0.003)	(0.003)	(0.004)
Ave. achievement	0.057**	-0.020*	-0.005
	(0.005)	(0.007)	(0.008)
Supportive principal	0.808**	-0.064*	-0.095**
	(0.017)	(0.024)	(0.029)
Blue bus	0.013	-0.013	0.008
	(0.014)	(0.020)	(0.024)

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

ONLINE APPENDIX TABLE 23—GRADE-LEVEL HETEROGENEITY IN COMPENSATION PREFERENCES

	Linear Probability		
	Elementary School	Middle School	High School
	(1)	(2)	(3)
Starting salary	0.090**	0.002	0.001
	(0.003)	(0.004)	(0.005)
Salary growth	0.193**	0.004	-0.007
	(0.012)	(0.011)	(0.013)
Bonus amount	0.035**	-0.001	-0.017*
	(0.006)	(0.008)	(0.008)
VAM only	-0.075**	0.010	0.011
	(0.019)	(0.018)	(0.020)
Replacement	0.014**	0.000	-0.000
	(0.001)	(0.001)	(0.001)
401k-style	0.079**	-0.010	0.011
	(0.015)	(0.021)	(0.022)
Premium (yearly)	-0.062*	0.010	-0.070
	(0.025)	(0.038)	0.039
Deductible	-0.306	-0.026	0.053
	(0.271)	(0.179)	(0.193)

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

ONLINE APPENDIX TABLE 24— GRADE-LEVEL HETEROGENEITY IN WORKING-CONDITION PREFERENCES

	Linear Probability		
	Elementary School	Middle School	High School
	(1)	(2)	(3)
Probationary period	-0.042**	-0.011	-0.012
	(0.005)	(0.007)	(0.007)
Term length	0.001	-0.004	-0.004
	(0.006)	(0.007)	(0.007)
Commute time	-0.003**	-0.002	-0.002
	(0.001)	(0.001)	(0.001)
Class size	-0.063**	0.012**	0.018**
	(0.002)	(0.002)	(0.002)
Assistance	0.022**	0.001	-0.003
	(0.001)	(0.002)	(0.002)

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

ONLINE APPENDIX TABLE 25— GRADE-LEVEL HETEROGENEITY IN STUDENT AND PRINCIPAL PREFERENCES

	Linear Probability		
	Elementary School	Middle School	High School
	(1)	(2)	(3)
Percent low income	-0.029**	-0.004	-0.008
	(0.005)	(0.007)	(0.007)
Percent minority	0.000	0.006	0.008
	(0.003)	(0.004)	(0.004)
Ave. achievement	0.037**	0.005	0.015
	(0.006)	(0.008)	(0.009)
Supportive principal	0.756**	0.035	0.015
	(0.021)	(0.030)	(0.032)
Blue bus	0.022	-0.010	-0.055*
	(0.018)	(0.025)	(0.027)

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

ONLINE APPENDIX TABLE 26—COMPENSATION ESTIMATES WITH DIMINISHING MARGINAL RETURNS FOR SIMULATION EXERCISES

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

Premium	-0.0821**	-0.1000**
	(0.0141)	(0.0160)
Observations	0.193	0.195
R-squared	31,820	31,820

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

ONLINE APPENDIX TABLE 27—WORKING CONDITIONS ESTIMATES WITH DIMINISHING MARGINAL RETURNS FOR SIMULATION EXERCISES

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

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

Online Appendix Table 28— Simulated Compensation Structure under Various Objectives without Class-Size Constraint

	Status quo	**Teacher-utility optimal**	**Teacher-retention optimal**	**Student-achievement optimal**
	(1)	**(2)**	**(3)**	**(4)**
Starting salary	$50,000	$140,232	$149,692	$110,973
Salary growth	1.80%	0.00%	0.00%	0.00%
Merit pay	$0	$1,464	$1,462	$5,000
VA-only merit	0	0	0	0
Replacement rate	69.0%	30.3%	26.3%	42.0%
Defined contribution	0	1	1	1
Insurance subsidy	$3,960	$0	$0	$0
Class size	28.7	52.4	55.4	42.9
Teacher utility	$79,200	**$116,500**	$116,800	$105,100
Teacher experience	9.03 years	13.8 years	**13.8 years**	12.2 years
Student achievement	0.092σ	0.138σ	0.135σ	**0.193σ**

ONLINE APPENDIX TABLE 29—ESTIMATED EFFECTS OF COMPENSATION ON RETENTION

	Transfer Districts	Depart Profession
	(1)	(2)
Salary_{t-4}	-0.003	-0.008
	(0.009)	(0.013)
Salary_{t-3}	-0.002	0.006
	(0.010)	(0.013)
Salary_{t-2}	0.011	0.006
	(0.010)	(0.013)
Salary_{t-1}	-0.025*	-0.041*
	(0.010)	(0.018)
Salary_{t}	0.025*	-0.074**
	(0.012)	(0.024)
Salary_{t+1}	-0.028*	-0.009
	(0.009)	(0.013)
Salary_{t+2}	-0.010	-0.010
	(0.009)	(0.012)
Salary_{t+3}	0.012	-0.034*
	(0.018)	(0.011)
Salary_{t+4}	-0.014	-0.015
	(0.008)	(0.010)
District-year FE	X	X
Experience-district FE	X	X
Experience-year FE	X	X
Mean outcome	5.72	9.29
Adjusted R-squared	0.048	0.048
Observations	1,556,106	1,557,106

*Note: * $p < 0.05$, ** $p<0.01$, *** $p < 0.001$.* This table presents the relationship between teacher departures and a distributed lag model of the salary schedule. The outcome variable in column (1) is teacher departures to other school districts. The outcome variable in column (2) is teacher departures from the profession. Data from the Texas Education Agency; standard errors clustered by teacher.

ONLINE APPENDIX TABLE 30—RELATIONSHIP BETWEEN UNION INFLUENCE AND BENEFIT SHARE

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

*Note: * $p < 0.05$, ** $p < 0.001$.* This table presents the relationship between union strength and the share of a teacher's compensation received in benefits. Data from LEFS; standard errors clustered at the state level. Union strength is measured uniformly on a scale from one to five describing what quintile of union strength each state has. For context, the mean dependent variable (benefit share) is 0.355, so a one unit change in union power is associated with an eight percent increase in benefit share.

Online Appendix H: Invitation Materials

EXHIBIT 1—EMAIL INVITATION

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

Dear [David],

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

Follow this link to the Survey: Take the Survey

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

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

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

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

Thanks so much!
Andrew

Follow this link to the Survey:
Take the Survey

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

Follow the link to opt out of future emails:

Click here to unsubscribe

EXHIBIT 2—OVERVIEW OF TEACHER SURVEY

Overview of Teacher Survey

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

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

How will confidentiality be maintained and your privacy be protected?

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

What should you do if you have questions?

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

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

EXHIBIT 3—PRE-QUESTION INSTRUCTIONS

The first 11 questions will ask you to choose between two hypothetical job offers.

We thank you for carefully considering each offer and designating which offer you would prefer.