

UC Berkeley

Working Paper Series

Title

Teacher Quality Policy When Supply Matters

Permalink

<https://escholarship.org/uc/item/81q0f4bc>

Journal

American Economic Review, 105(1)

Author

Rothstein, Jesse

Publication Date

2012-09-01

Data Availability

The data associated with this publication are available at: <http://irle.berkeley.edu/teacher-quality-policy-when-supply-matters/>

Peer reviewed

Teacher Quality Policy When Supply Matters

Jesse Rothstein*

University of California, Berkeley and NBER

September 2012

Abstract

Recent proposals would strengthen the dependence of teacher pay and retention on performance, in order to attract those who will be effective teachers and repel those who will not. I model the teacher labor market, incorporating dynamic self-selection, noisy performance measurement, and Bayesian learning. Simulations indicate that labor market interactions are important to the evaluation of alternative teacher contracts. Typical bonus policies have very small effects on selection. Firing policies can have larger effects, if accompanied by substantial salary increases. However, misalignment between productivity and measured performance nearly eliminates the benefits while preserving most of the costs.

*E-mail: rothstein@berkeley.edu. I thank Sarena Goodman for excellent research assistance and David Card, Sean Corcoran, Richard Rothstein, Cecilia Rouse, and conference and seminar participants at APPAM, Berkeley, IEB, IRP, NBER, Northwestern, and NYU for helpful discussions. I am grateful to the Institute for Research on Labor and Employment at UC Berkeley for research funding.

1 Introduction

Recent education policy debates have centered on teacher quality.¹ Secretary of Education Arne Duncan lays out the agenda: “We have to reward excellence....We also have to make it easier to get rid of teachers when learning isn’t happening” (Hiatt, 2009).² Researchers have focused on developing and validating measures of teacher effectiveness (Chetty et al., 2011; Bill & Melinda Gates Foundation, 2012), though questions remain (Rothstein, 2011; Corcoran, 2010). By contrast, relatively little attention has been paid to the design of policies that will use the new measures to improve educational outcomes.

Several recent experiments have examined the short-term effects of performance bonuses, with generally disappointing results (Goodman and Turner, Forthcoming; Fryer, 2011; Springer et al., 2010; though see Fryer et al., 2012).³ These studies were designed to detect teacher effort responses, which may be the wrong margin. Many observers believe that variation in teacher effectiveness primarily reflects largely immutable personality traits.⁴ Under this view, the primary mechanism by which instructional quality might be improved is through selection.

¹In a 2010 manifesto, sixteen big-city school superintendents confidently state that “the single most important factor determining whether students succeed in school...is the quality of their teacher” (Klein et al., 2010). Influential advocates promise that policies aimed at improving teacher quality can “turn our schools around” (Gates, 2011).

²Secretary Duncan’s suggestions are not the only potential routes to improved instruction. Alternatives include improved selection on entry into the profession or more or better professional development. Researchers have had trouble identifying characteristics observable at the time of hiring that are strongly correlated with subsequent effectiveness (Hanushek and Rivkin, 2006; Clotfelter et al., 2007; Rockoff et al., 2011). Taylor and Tyler’s (Forthcoming) examination of a formative evaluation program for experienced teachers found large impacts on teachers’ subsequent performance.

³The evidence from the developing world is more positive. See, e.g., Lavy (2002) and Duflo et al. (2012); but also see Glewwe et al. (2010).

⁴Klein et al. (2010), for example, urge us to “stop pretending that everyone who goes into the classroom has the ability and temperament” to be an effective teacher.

A well designed contract could make the profession more attractive to effective teachers and less attractive (or perhaps unavailable) to ineffective teachers (Lazear, 2003). We know little about effects of this type. Career decisions depend on expected compensation many years in the future, and short-term experimental interventions cannot have large effects on this. Even quasi-experimental approaches are not promising. Performance pay systems have generally been short-lived (Murnane and Cohen, 1986), so potential teachers are unlikely to expect that recent policy experiments will persist for very long.

In this paper, I use simulations to examine the selection effects of alternative teacher contracts. I develop a stylized model of the teacher labor market that incorporates heterogeneity in teacher ability. Teacher supply responses derive from a dynamic discrete choice model in which graduates and experienced teachers choose between teaching and alternative occupations on the basis of the compensation on offer.⁵ Decisions to enter into teaching depend on risk-adjusted expected compensation over the whole career. Similarly, experienced teachers consider their expected compensation over the remainder of their potential careers in deciding whether to exit the profession. Alternative contracts affect the future compensation and security of a teaching job, with differential anticipated impacts on teachers who vary in their estimates of their own ability. These impacts in turn influence both decisions to enter the profession and to exit for other opportunities.

Two consensus results from the recent teacher quality literature are that teacher ability is difficult or impossible to measure directly (via, e.g., education or

⁵Dynamic occupation choice models include Adda et al. (2011) and Keane and Wolpin (1997); see also Wiswall (2007) on teacher licensing requirements. Essentially static models of teacher attrition include Murnane and Olsen (1989) and Dolton and van der Klaauw (1999). Tincani (2011) develops a static model of teacher sectoral choice under policies like those considered here.

personality characteristics) and that realized performance can provide a noisy but informative measure. Thus, I assume that compensation and retention decisions can condition only on a sequence of noisy performance signals – which might be “value added” scores or some alternative – and not directly on teacher ability. Prospective teachers start with limited information about their own abilities, and then update their estimates with each performance measure. A teacher who receives positive signals concludes that she is likely to receive an above average number of performance bonuses in future years or to have a below average probability of being fired for poor performance, while a teacher who receives negative signals concludes the opposite. These expectations drive the teacher’s dynamic decision-making about whether to enter the profession and, having entered, to remain.

Given the extremely limited variation in extant teacher contracts, I do not attempt to estimate the model directly. Instead, I calibrate it using parameters obtained from estimates in the literature. I attempt to choose parameters to make the best realistic case for performance-based contracts, and explore the robustness of the results to specific parameter values through extensive sensitivity checks.

My policy analysis is closely related to studies of teacher retention and non-retention by, e.g., Staiger and Rockoff (2010) and Boyd et al. (2011). These studies ignore behavioral responses. In Staiger and Rockoff’s (2010) simulation of teacher tenure policies, for example, the district can draw without limit from the current applicant pool to replace poor performers who are fired, without raising salaries. Not surprisingly, then, the optimal tenure denial rate is extremely high. My model augments Staiger and Rockoff’s framework with a non-degenerate teacher labor market. An increased firing rate requires higher salaries, both to compensate prospec-

tive teachers for the increased risk and to attract enough additional applicants to replace fired teachers.⁶ I find that the required salary increase is substantial, even for much much lower firing rates than those considered by Staiger and Rockoff (2010). Nevertheless, the results indicate that firing policies can be cost effective, albeit with smaller benefits than have sometimes been promised.

My framework allows me to consider a broader class of contracts than do Staiger and Rockoff (2010). I focus on a performance bonus based on recent outcomes and a policy of ongoing retention decisions. In the appendix, I present additional results for a traditional up-or-out tenure decision after the second year of the career (as in Staiger and Rockoff, 2010) and a performance pay system in which pay is continuously related to demonstrated effectiveness.

In order to focus on the selection effects of performance accountability policies, I rule out by assumption any other effects of these policies. Effort is irrelevant, and teachers can do nothing to influence their actual or measured performance. This neglects the important possibility that high-stakes accountability could lead to distortion of the performance measure (Campbell, 1979), perhaps through narrowing of curricula and redirection of effort toward measured outcomes (Holmstrom and Milgrom, 1991; Glewwe et al., 2010), changes in student assignments (Rothstein, 2009), or even outright cheating (Jacob and Levitt, 2003). It could also crowd out intrinsic motivation, thereby *lowering* teacher effort (Jacobson, 1995), or discour-

⁶Although there is considerable slack in the teacher labor market now, following substantial layoffs during the Great Recession, as recently as 2007 education policymakers worried about where they would find enough qualified new teachers to fill the expected openings (Chandler, 2007; Gordon et al., 2006). Shortfalls are traditionally filled by hiring teachers with “emergency” credentials, often semi-permanently. Some have hypothesized that current credentialing rules are unrelated to quality. If so, loosening of requirements could serve to offset other changes that reduce supply. But changes in entry requirements need not be accompanied by changes in pay or retention policies, so the two policies can be evaluated independently. I focus on the latter.

age cooperation among coworkers. The potential for manipulation and, worse, goal distortion militates against high-stakes uses of the performance measures (Baker, 1992, 2002). In Section 5 I extend the model to allow for an imperfect alignment between true productivity and the output that is measured and for possible distorting effects of incentives. I find that these considerations are extremely important – even a very limited ability of teachers to manipulate their measured performance can undo nearly all of the positive effects of a firing policy.

2 What are the Policies of Interest?

Rick Hess, responding to the POINT study of teacher performance bonuses (Springer et al., 2010), argues that identifying the selection effects of such policies requires that the researcher “start by identifying a couple thousand high school students, follow them for fifteen or twenty years, and study whether alterations to the compensation structure of teaching impacted who entered teaching, how they fared, and how it changed their career trajectory” (Hess, 2010). Even if this could be accomplished, the study “wouldn’t tell us what to do today [and] wouldn’t generate much in the way of findings until the 2020s” (Hess, 2010). Efforts to evaluate selection effects via natural experiments face similar challenges (Murnane and Cohen, 1986).

This motivates my structural modeling strategy, as such an approach can predict the effects of policies that have not yet been implemented – if the model is correctly specified. In principle, the parameters of the model could be identified using data on teachers subject to traditional contracts. But because these contracts typically involve a “single salary schedule” that is totally invariant to teacher effectiveness, results would be highly dependent on functional form and distributional

assumptions. I thus calibrate the model rather than estimating it, relying on the best evidence from the literature about the various parameter values.

There are two broad components of the model: The performance measurement system and the teacher labor market. A great deal of recent research has examined the former topic, focused around “value added” models that measure a teacher’s effectiveness based on her students’ test score growth. I use estimates from this literature to calibrate the performance measurement component of the model. Nothing in the model is specific to a value-added-based system, however; the model could equally well describe contracts based on more traditional, classroom-observation-based, performance measures.

The second major component is the interaction between teacher accountability policies and the teacher labor market. Calibration of this component is much less straightforward. I discuss here several aspects of the policies of interest that bear on the teacher labor market.

First, I focus on modeling policy changes implemented at the level of the state or nation rather than by individual districts. This implies smaller labor supply elasticities than would apply to a small district (Manning, 2005).⁷ Relatedly, I rule out the “dance of the lemons,” whereby a teacher fired by one district is rehired by another; in my model, fired teachers must exit the profession.

Second, an important issue in my analysis is uncertainty about a teacher’s ability that is gradually resolved through her demonstrated performance on the job.

⁷Lazear’s (2000) famous Safelite Auto Glass study examines a firm-level performance pay program; a similar program implemented at the industry level would almost certainly have smaller selection effects. The ongoing official evaluation of the federal Teacher Incentive Fund will assign schools to treatments within participating districts (Mathematica Policy Research, undated), severely limiting its ability to identify policy-relevant effects on teacher recruitment and retention.

Accordingly, I distinguish between decisions to enter the teaching profession and those about exiting later. Because uncertainty is greatest at the beginning of the career, entry decisions are relatively insensitive to performance-dependent contracts, while exit decisions become gradually more sensitive as the career goes on.

Third, as suggested above by Hess (2010), occupation choices depend on both current salary offers and on anticipated future salaries.⁸ My model explicitly incorporates forward-looking expectations and uncertainty coming from limited information about one's own ability and from noise in the performance measurement system. I rule out, however, uncertainty about the future direction of policy: I assume that everyone assumes that any proposed contract will be unchanged for the duration of all current teachers' careers, and I examine steady-state effects after all teachers recruited under a prior contract have retired.

Finally, teacher quality depends on demand as well as on supply. I assume that districts are unable to distinguish teacher ability at the point of hiring.⁹ Their only options for managing quality are to adjust contract parameters to induce self selection on the part of potential teachers and/or to retain teachers selectively after observed performance reveals a teacher's quality. The district obtains no value from excess supply, so I assume that it adjusts base wages to the point that total labor supply (net of that offered by teachers who the district chooses to fire) matches that obtained under the baseline single salary schedule. This amounts to treating

⁸I am not aware of good estimates of the return to teaching experience in non-teaching jobs. If experienced teachers are paid in other occupations like inexperienced workers, policies that raise the turnover rate may dramatically lower the expected lifetime returns to entering teaching.

⁹Ballou (1996) hypothesizes that salaries are weakly linked to quality in the cross section because districts facing excess supply do a poor job of selecting the highest-quality applicants. This claim is not inconsistent with a supply-side effect of salaries on the quality of applicants, as implied by evidence that teacher quality has declined as high-ability women's non-teaching options improved (Corcoran et al., 2004).

the district’s labor demand as totally inelastic, consistent with laws and collective bargaining contracts that commonly specify class sizes.

3 The Model

I develop the model in several parts. Section 3.1 defines the performance measure and the Bayesian learning process. Section 3.2 discusses entry and exit decisions, which depend on both the contract terms and the teacher type. These are motivated by an on-the-job search model. Finally, Section 3.3 describes the performance-linked contracts that I consider. Additional contracts are considered in the appendix.

3.1 Effectiveness, Performance Measurement, and Learning

Individual i has fixed ability τ_i as a teacher. In the current pool of teachers ability is normally distributed with mean 0 and standard deviation σ_τ , though new contracts may change the selection process and thereby alter that distribution.

A teacher’s output depends on her ability and her experience, t , with known return-to-experience function $r(t)$: $y_{it}^* = \tau_i + r(t)$. Each year, a noisy productivity measure is observed by both the teacher and the employer:¹⁰

$$y_{it} = \tau_i + \varepsilon_{it}. \tag{1}$$

The noise component, ε , is i.i.d. Gaussian with mean 0 and standard deviation σ_ε .

The performance measure is unbiased – all teachers draw their ε s from the same

¹⁰In practice, the observed signal is $y_{it}^* + \varepsilon_{it} = y_{it} + r(t)$. But so long as the $r(\cdot)$ function is known, all parties can easily back out y_{it} .

distribution, regardless of who they teach or the methods they use.¹¹

Prospective teachers have only limited information about their τ_i s. At entry, teacher i 's prior is $\tau_i \sim \mathbb{N}(\mu_i, \sigma_\tau^2 - \sigma_\mu^2)$, where μ_i represents the teacher's private information and $\mu_i \sim \mathbb{N}(0, \sigma_\mu^2)$ in the population of current teachers.¹² The precision of potential teachers' information can be summarized by $h \equiv V(E[\tau|\mu])/V(\tau) = \sigma_\mu^2/\sigma_\tau^2$, where $h = 1$ corresponds to perfect accuracy and $h = 0$ to a total lack of information. The employer cannot observe μ_i or τ_i , and compensation and retention thus depend only on the sequence of y_{it} s.

Incumbent teachers update their priors rationally as performance signals arrive. The teacher's posterior after t years is

$$\tau | \theta_t \sim \mathbb{N}\left(\frac{t^{-1}\sigma_\varepsilon^2\mu + (1-h)\sigma_\tau^2\bar{y}_t}{t^{-1}\sigma_\varepsilon^2 + (1-h)\sigma_\tau^2}, \frac{1}{t\sigma_\varepsilon^{-2} + (1-h)^{-1}\sigma_\tau^{-2}}\right), \quad (2)$$

where $\bar{y}_t \equiv t^{-1}\sum_{s=1}^t y_s$ is the average performance signal to date. I denote the posterior mean – the first term in (2) – by $\hat{\tau}_{it}$. As t gets large, the influence of the original guess shrinks, and $\hat{\tau}_{it}$ converges toward the true ability τ_i .

3.2 The Teacher Labor Market: Entry and Persistence

Prospective teachers have von Neumann-Morgenstern utility $u(w)$, where w is annual compensation, and discount rate δ .¹³ Each prospective and incumbent teacher

¹¹It is not clear whether real-world performance measures have this property – see Rothstein (2010) and Chetty et al. (2011).

¹²If low-ability prospective teachers overestimate their own effectiveness, the effect of performance incentives on recruitment would be diluted – even bad teachers would respond to incentives meant for good ones.

¹³To my knowledge, this is the first dynamic occupational choice model to allow for risk aversion (i.e., concave $u(\cdot)$). A more complete model would define $u(\cdot)$ over annual consumption and allow agents to borrow and save to smooth consumption over time. I do not pursue this here. I discuss an

draws a single outside job offer at the beginning of each year t , $1 \leq t \leq T$. These offers are summarized by their continuation values, ω_t .

A prospective teacher forecasts the utility she will obtain from a teaching career, which I denote $V_1(\mu; C)$ to emphasize that this may depend both on her private information μ and on the terms of the contract C , and enters teaching if $V_1(\mu; C) > \omega_1$. An incumbent teacher's forecast of her remaining utility in teaching, $V_t(\theta_{t-1}; C)$, depends on the state variable $\theta_{t-1} \equiv \{\mu, y_1, \dots, y_{t-1}\}$ and on C . If $V_t(\theta_{t-1}; C) < \omega_t$, the teacher accepts the outside offer. Teachers who accept outside offers, either initially or later, can not reenter teaching.

After each year of teaching, the teacher receives a new performance signal, and uses this to better forecast her inside earnings. She also receives a new draw from the outside offer distribution. Thus, V_t is defined recursively:

$$V_t(\theta_{t-1}; C) = E[u(w_t) + \delta \max(\omega_{t+1}, V_{t+1}(\theta_t; C)) | \theta_{t-1}]. \quad (3)$$

The expectation is taken over the teacher's posterior τ distribution following period $t - 1$, as given by (2), and over the distribution of the noise term ε_{it} . Careers end after T periods, so $V_T(\theta_{T-1}; C) = E[u(w_T) | \theta_{T-1}]$.

ω_t is assumed independent of θ_{t-1} and τ . This is consistent with most of the available evidence, which generally indicates little relationship between teaching effectiveness and characteristics observable on entry (Hanushek and Rivkin, 2006; Rockoff et al., 2011). Several studies find that effectiveness is negatively correlated with the probability of exit from teaching (Krieg, 2006; Goldhaber et al., 2011); given the weak or nonexistent returns to effectiveness in teaching, one alternative model that captures some income smoothing in Section 4.3.

would expect the opposite if teaching ability were correlated with outside wages. I weaken this assumption later.

The ω_t distribution is calibrated so that the annual exit hazard under the base contract – the probability that ω_t exceeds $V_t(\theta; C_0)$, where C_0 is the baseline, single-salary contract discussed below – is λ_0 and the elasticities of entry and exit with respect to certain, permanent changes in w are η and $-\zeta$, respectively.¹⁴ The Appendix discusses the distributional assumptions that generate this.

Equation (3) does not have a closed-form solution, so I evaluate the value function numerically using an algorithm described in the Appendix. To simulate the impact of an alternative contract C , I draw teachers from the joint distribution of $\{\mu, \tau\}$, then draw performance measures $\{y_1, \dots, y_T\}$ for each. For each teacher, I compute V_t at each year t under contract C , and use this to compute the effect of contract C on the probability of entering the profession and, conditional on entering, on surviving to year t . Note that this does not require modeling the distribution of $\{\mu, \tau\}$ in the *population* of potential teachers – under my constant elasticity assumptions, changes in the returns to teaching induce proportional changes in the amount of labor supplied to teaching by each type that do not depend on the number of people of that type in the population.

An important parameter governing the effect of alternative contracts is the cost to a teacher of being fired. I assume that a teacher who is fired after year t receives continuation value equal to $(1 - \kappa)$ times the continuation value obtained under contract C_0 (under which no one is ever fired). This captures the empirical

¹⁴As $T \rightarrow \infty$ the average career length approaches the inverse of the annual exit hazard, and the elasticity of the career length with respect to the inside wage converges toward ζ . With the parameters used below ($T = 30$, $\lambda_0 = 0.08$, and $\zeta = 3$), the elasticity is roughly $0.77\zeta = 2.3$. The total labor supply elasticity is the sum of the entry and career length elasticities, $\eta + 0.77\zeta$.

fact that workers who lose their jobs see long-run earnings declines (von Wachter et al., 2009). The firing penalty κ is not paid by someone who voluntarily exits in advance of an expected termination. Thus, if κ is large, teachers who anticipate a high probability that they will eventually be fired will voluntarily exit at high rates.

3.3 Teacher contracts

I treat as the baseline a single salary contract C_0 under which all teachers are retained every year (though they may depart voluntarily) and pay rises with t but is independent of θ : $w_{it}^0 = w^0(1 + g(t))$, with $g'(\cdot) \geq 0$. Alternative contracts base either the compensation or the retention decision on the sequence of performance signals to date. I consider two alternatives, performance-based bonuses and performance-based non-retention (i.e., firing):

Bonus Bonuses are awarded to teachers with sufficiently high measured performance, averaged across two years to reduce the influence of noise. Thus, in year t bonuses are awarded to all teachers with $\frac{y_{it} + y_{i,t-1}}{2}$ above a threshold y^B ; first-year teachers are ineligible. y^B is calibrated based on the baseline distribution of $\frac{y_{it} + y_{i,t-1}}{2}$ to ensure that in the absence of behavioral responses a fraction f^B of teachers would receive bonuses each year. In order to balance the increased labor supply induced by the bonus program, base salaries are reduced to a fraction $\alpha^B < 1$ of the baseline salary w_{it}^0 .¹⁵ Thus, compensation is $w_{it}^B = \alpha^B w_{it}^0 (1 + b * e_{it})$, where e_{it} is an indicator for bonus receipt and b

¹⁵Bonus programs have often been implemented as add-ons to the preexisting base salary. In my framework, this produces excess labor supply, and districts have no choice but to choose randomly from the pool of applicants. The productivity benefits are the same as with a lower base salary, but costs are much higher.

indexes the size of the bonus (as a share of base pay).

Firing In this contract, firing decisions are made on an ongoing basis – any teacher for whom the district’s posterior mean falls below a threshold y^F is dismissed.¹⁶ Thus, a teacher who performs extremely badly in the first year or two is fired right away, while a teacher who squeaks above that threshold but continues to perform badly is fired somewhat later. y^F is calibrated so that the share of teachers in the current ability-experience-performance history distribution whose posterior means fall below the threshold is f^F . Pay for teachers who are not fired is according to the single salary schedule, $w_{it}^F = \alpha^F w_{it}^0$, with $\alpha^F > 1$ to ensure adequate labor supply.

The firing contract differs from the tenure rules considered by Staiger and Rockoff (2010), under which the district is forced to make a once-and-for-all retention decision early in the teacher’s career. Tenure contracts are more common than the one I consider, but make poor use of the available information – there is no reason to ignore post-tenure performance information entirely.

In the appendix, I consider a more traditional tenure contract. I also consider there a contract that allows pay to depend continuously on performance, with a larger variable component for experienced teachers for whom the district has better information. Neither these nor the contracts above are optimal. In my model, where effort is irrelevant, the optimal pay schedule would have low or zero annual pay and a very large performance-dependent retirement bonus. The firing rule would have

¹⁶I assume the district uses a prior for entering teachers of $\tau \sim \mathbb{N}(0, \sigma_\tau^2)$, resulting in posterior mean $\frac{\sigma_\tau^2}{(t-1)\sigma_\epsilon^2 + \sigma_\tau^2} \bar{y}_{i,t-1}$. This prior is correct under the baseline contract, but under the firing contract the τ distribution for entering teachers will differ from this. I ignore this complication.

a threshold that rises even faster with t than in my firing contract, in recognition of the option value of retaining a teacher for whom the prior variance is high.

3.4 Calibration

Table 1 lists the key parameters of the model along with the values that I use. I discuss the choice of baseline values in each category in turn.

I calibrate the effectiveness and measurement parameters from the value-added literature. The standard deviation of teacher value-added for students' end-of-year test scores has been widely estimated to be between 0.1 and 0.2, with 0.15 as a reasonable central estimate (e.g., Rivkin et al., 2005; Rothstein, 2010; Chetty et al., 2011). The same research typically shows important experience effects in the early years of the career that level off later; the specific value for $r(t)$ in Table 1 is drawn from Staiger and Rockoff's (2010) estimates for New York City. A number of papers also examine the year-to-year correlation of value-added measures (Sass, 2008; Bill & Melinda Gates Foundation, 2010). My chosen value for σ_ε corresponds to a reliability ratio for y (defined as $\frac{V(\tau)}{V(y)} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2}$) of 0.4, at the upper end of the range surveyed by Sass (2008).

A number of studies have found that observable teacher characteristics are poor predictors of future effectiveness. Rockoff et al. (2011) are among the most successful at predicting future value-added, using a number of academic and personality characteristics, but obtain an h of only 0.1. Of course, teachers may have more information about their own personalities than can be captured by the Rockoff et al. (2011) survey.¹⁷ Nevertheless, my assumption that $h = 0.25$ most likely

¹⁷Participants in the POINT study (Springer et al., 2010) were asked to forecast their probability

overstates the true value.

There are few good estimates of the key teacher preference parameters. The discount rate is relatively standard. Zero risk aversion is unlikely; one might expect teachers – who have self-selected into a very secure but low paying occupation – to be unusually averse to risk (Flyer and Rosen, 1997). I use linear utility as a baseline, and explore risk aversion in Section 4.3.

The exit elasticity, ζ , is taken from Ransom and Sims' (2010) study of salary variation across Missouri school districts. This study focuses on exit to other school districts, and thus most likely overstates the elasticity of exit from the profession. I arbitrarily assume that the entry elasticity η is the same (in absolute value). I assume that the annual exit hazard under the baseline single salary contract is constant at 8%, and that careers end after $T = 30$ years. This is roughly consistent with the observed national data, though in these data exit rates are somewhat higher in the first years of teachers' careers; see Appendix Figure 1.

A teacher who is fired obtains a continuation value $\kappa = 10\%$ below what is obtained under the baseline contract. This is likely an understatement – von Wachter et al. (2009) find that workers displaced by mass layoffs see their earnings decline by 20-30% relative to a control group, with effects that persist for at least 20 years.¹⁸

I assume that the single salary contract provides for a 1.5% (real) increase of winning a performance award. These forecasts — from experienced teachers who certainly knew more about their own effectiveness than would an entering teacher — were uncorrelated with actual award receipt. This suggests that h is quite small.

¹⁸The effects of mass layoffs may overstate the effect if layoffs are concentrated in declining occupations or industries. On the other hand, even in declining sectors at least some laid off workers are able to find reemployment in the same sector; I assume that a fired teacher must exit the occupation. Moreover, future employers may react more negatively to learning that a job candidate was fired from a previous position for poor performance than that she lost her job in a mass layoff.

for each year of experience. The alternative contracts are based on this contract and thus provide similar average experience premia. The bonus contract provides a 20% bonus for teachers whose two-year moving average performance exceeds a fixed threshold $y^B = 0.178$, set to ensure that $f^B = 25\%$ of the current teaching workforce would get bonuses.

The firing contract is calibrated so that a teacher is fired whenever the district's posterior mean for her ability falls below $y^F = -0.159$. This threshold is chosen to ensure that $f^F = 10\%$ of current teachers would be fired in the first year of the contract's implementation, though the steady state firing rate would be much lower. Given the other parameter values, a new teacher would need $y_1 < -0.40$ to be fired after one year, $\frac{1}{2}(y_1 + y_2) < -0.29$ to be fired after two years, and $\frac{1}{3}(y_1 + y_2 + y_3) < -0.24$ to be fired after three. Under the current τ distribution, less than 5% of teachers would be fired after the first year, a slightly larger share of those who remain would be fired in the second year, and firing rates would decline thereafter; 22% of teachers would be fired before the end of a 30-year career.

Both y^B and y^F are fixed – if the alternative contracts attract more high- τ teachers then more bonuses would be paid or fewer teachers would be fired.

The final parameters are α^B and α^F , the adjustments to base pay under the bonus and firing contracts. These are chosen, given the other parameters, to ensure the same total number of teachers (in steady state) as are obtained under the baseline contract. This requires a 3.6% reduction in base pay under the bonus contract and a 5.4% increase under the firing contract.

4 Results

4.1 Noise, information, and incentives

The incentive faced by a teacher i with ability prior $\hat{\tau}_{it}$ depends on the link between this prior and her true ability, the link from true ability to the performance signal, and the link from the performance signal to the contract terms. Moreover, the success of a contract depends on the average incentive perceived by teachers at each true ability level τ_i , among whom there may be much variation in $\hat{\tau}_{it}$.

Under the bonus contract, the rule of iterated expectations can be used to express the subjective probability in year t of winning a bonus in year $s > t + 1$ (ensuring that y_t does not directly affect bonus eligibility) as:

$$E[e_{is} | \hat{\tau}_{it}] = E \left[E \left[E[e_{is} | y_{is}, y_{i,s-1}] | \tau_i \right] | \hat{\tau}_{it} \right]. \quad (4)$$

I am able to omit the outer conditioning variables from the inner expectations because the inner variables capture all of the relevant information – bonus receipt is independent of ability conditional on performance over two years, and performance is independent of perceived ability conditional on true ability.

The innermost term is straightforward – e_{is} is merely an indicator for $\frac{y_{is} + y_{i,s-1}}{2} > y^B$. But each of the two outer expectations serves to smooth out the incentives. The dotted line in Figure 1 plots the probability that a teacher at each ability (τ) level will win a bonus in a given year, $E[e_{is} | \tau_i]$.¹⁹ The other series show the effects of

¹⁹In this figure and in those that follow, I express ability in terms of the percentile rank within the *current* teacher distribution (which, recall, is Gaussian with mean 0 and standard deviation $\sigma_\tau = 0.15$). Of course, under alternative policies this distribution would change. The fixed-norm percentile scores are simply a convenient scale.

shortening or lengthening the averaging period. They show that using only one year of measured performance produces extremely high rates of misclassification, with bonus probabilities above 10% for all teachers in the upper two-thirds of the ability distribution. Averaging over five years reduces the probability that a bottom-half teacher will win a bonus to less than 2%, but still yields substantial error rates for teachers near the threshold: While top decile teachers win bonuses an average of 80% of the time, those in the next decile win with only 40% probability. The two-year bonus contract is intermediate between these, with meaningful incentives for top-third teachers but only limited distinctions among them.

These incentives are further attenuated by uncertainty about one's own ability. Figure 2 shows $E[e_{is} | \hat{\tau}_{it}] = E[E[e_{is} | \tau_i] | \hat{\tau}_{it}]$ at different points in the career. For early-career teachers, the $\hat{\tau}_{it}$ distribution is quite compressed. Thus, even a teacher at the 90th percentile of the μ distribution thinks she has only a 37% chance of receiving a bonus in any given year of her career. As teachers accumulate information, they quickly learn their places in the distribution. After one year, the teacher at the 90th percentile of the posterior mean distribution thinks her chance of receiving a bonus is 42%, and this rises to 45% after two years and 49% after 5 years. By that point, teachers' posteriors are fairly tight, and the curve in Figure 2 closely resembles that in Figure 1.

Figure 2 illustrates the allocation of incentives by teachers' perceived ability, $\hat{\tau}_{it}$. But the key question for the efficacy of the bonus system is whether teachers who actually are of high ability perceive their future pay to have risen. That is, the incentives that the teacher contract creates to attract good teachers are governed by $E[E[w_{is} | \hat{\tau}_{it}] | \tau_i]$, which is necessarily flatter than $E[w_{is} | \hat{\tau}_{it}]$ because be-

cause $E[\hat{\tau}_{it} | \tau_i]$ has a slope less than one. Figure 3 shows the average anticipated probability of winning a bonus by percentile of *true* ability. At entry there is very little differentiation except at the extreme tails of the distribution. But perceived incentives become much better targeted as teachers gain experience. Thus, while incentive effects of a bonus system are weak at the recruitment stage, later attrition decisions may be more sensitive to these incentives.

As Figure 1 illustrates, a big source of slippage in the bonus program is the use of only two years of performance data for determination of bonus eligibility, even when more are available. This suggests that the firing contract, which uses all available performance data for each year's retention decisions, may be more effective. I turn now to this contract. The solid line in Figure 4 shows the probability that a teacher of ability τ will be fired at some point over a 30 year career under this contract. Not surprisingly, the firing policy is much more accurate than is the bonus policy (compare to Figure 1): A teacher at the 10th percentile has a 93% chance of being fired, where a teacher at the 90th percentile had only a 54% chance of receiving a bonus in any given year, while a median teacher has a 9% chance of receiving a bonus but only a 4% chance of ever being fired.

As with the bonus policy, however, the incentives created by the firing contract are attenuated by teachers' uncertainty about their own abilities. The dashed line in Figure 4 shows the average subjective probability of ever being fired, measured at the beginning of the career and averaged across all prospective teachers at each ability level. It shows that there is relatively little difference between high and low ability prospective teachers in their subjective assessments of the likelihood that they will be recognized as ineffective.

There is a close, albeit imperfect, mapping from the subjective probabilities of positive and negative outcomes graphed in Figures 3 and as the dashed line in 4 to the average values of teachers of different abilities under the two contracts. Figure 5 shows average continuation values of teachers under the two contracts, by ability level and years of experience.²⁰ Because the V scale is not intuitive, I report equivalent variations: Changes in salaries under the single salary contract that would yield the same values. An estimate of (for example) +5% means that w^0 would need to rise by 5% under contract C_0 to yield the same value as is obtained under the alternative contract. The figure shows that the bonus contract produces the equivalent of a 1.3% salary increase for the average 95th percentile teacher at entry, and a 6.9% increase after five years.²¹ These small changes suggest that any self-selection responses to the bonus program will be quite modest, even with relatively large labor supply elasticities. The firing contract achieves a steeper slope, but even it creates only weak incentives for self selection: The range of continuation values is equivalent to only about 7% of salary differentiation at entry, growing to about 15% after two years and shrinking thereafter.

4.2 Impact of incentives

Figure 5 indicates that the performance pay contract creates quite modest incentives to encourage highly effective teachers to enter and remain in teaching, and that the

²⁰Averages are computed over all teachers who have not been fired to date, ignoring voluntary exit decisions. This means that the values shown for experienced, low-ability teachers under the firing contract reflect outlier teachers who have been unusually lucky in their measured performance.

²¹Recall that base salaries are reduced by an amount $1 - \alpha^B$ under the bonus contract. Thus, the average equivalent variation for the lowest ability teachers approaches $\alpha^B - 1 = -3.6\%$ as the (subjective) probability of future bonus receipt approaches zero.

firing contract creates somewhat larger but still not enormous incentives – though a potentially more important effect of this contract is that it forces many teachers to leave even though they would prefer to remain. What do these estimates imply for the recruitment and retention of teachers of different abilities?

I begin by examining recruitment. Figure 6 shows the number of entering teachers at each ability percentile under each contract, expressed as a percentage of the number obtained under the baseline contract. Both alternative contracts entice more high ability and fewer low ability teachers to enter teaching, with the firing contract much more successful than the bonus contract at the top end but much less successful at the bottom end. (Note that the firing contract requires many more recruits in total, to replace those from earlier cohorts who have been fired.)

Figure 7 shows average career length under the four contracts. The bonus contract has only small effects on this margin, concentrated at the very top of the ability distribution. To offset the increased labor supply of high-ability teachers, base salaries are reduced by $1 - \alpha^B = 3.6\%$ under this contract, reducing career lengths by about 5% for below-median teachers.

The firing contract has a more dramatic effect. At the bottom of the ability distribution, career lengths shorten dramatically, by as much as 85% for the very weakest teachers. This primarily reflects firing decisions. Substantial salary increases are required to offset this; I find that salaries under the firing contract must rise by $\alpha^F - 1 = 5.9\%$ to yield enough teachers. This reduces voluntary attrition for teachers who do not expect to be fired, lengthening the average career of above-median teachers by about 12%. The higher salaries also reduce voluntary attrition among low-ability teachers, but because these teachers are generally fired

quite early in their careers this has little effect.

Figure 8 presents the impact of the two contracts on the steady state number of teachers at each ability level, combining entry and career length effects. Not surprisingly, the bonus contract has relatively small effects, reducing the number of low ability teachers by about 10% and increasing the number of high ability teachers by a bit more, as much as 40% at the very top of the distribution. The firing policy is much more effective, attracting slightly fewer of the highest ability teachers but more than making up through this with larger increases in the number of average-to-high ability teachers and dramatic reductions at very low ability levels.

Table 2 shows the effects of the two contracts on teacher ability, experience, effectiveness (combining ability and experience effects), and salaries. Bonuses raise average ability only slightly, while the firing policy is nearly three times as effective. Neither policy has large effects on the experience distribution, so net effects on teacher productivity – +0.015 for bonuses and +0.040 for firing – are quite close to the gross effects on teacher ability.

The result that the firing policy has negligible effects on the number of inexperienced teachers contrasts sharply with Staiger and Rockoff's (2010) result that teacher experience effects constrain the scope of tenure policies (albeit at much higher firing rates). There are two explanations for this. First, in the Appendix I show that a fixed- t tenure policy has larger (but still modest effects) on the steady state number of inexperienced teachers than does the ongoing firing policy I consider here. Second, and much more important, is the role of labor market interactions: The firing policy must be accompanied by substantial salary increases, and the resulting reductions in turnover among experienced teachers largely offset the

increases due to firing of new teachers. Thus, the main tradeoff that this policy involves is not with experience but with salary costs, which rise by 5.9% under the firing policy. Salary costs also rise under the bonus policy, by 1.8%, as base salaries cannot be reduced enough to offset the cost of the bonuses themselves.

It is useful to compare the cost effectiveness of the two contracts to a more traditional way to use extra educational resources, for class size reduction. A calculation based on Krueger's (1999) analysis of the STAR class size experiment suggests that a 1% reduction in class size would cost about 3.0% of the teacher salary budget and would raise student achievement by about 0.004 standard deviations.²² Both of the contracts considered here are substantially more cost effective than this, at least in my stylized simulation. One implication is that it would be possible to pay for each program by raising class sizes, rather than by raising total expenditures, while still retaining positive student achievement effects.

Another implication is that the policies could be expanded while remaining cost effective. I focus here on the firing policy, as this is more naturally scaled. (Few have proposed making more than about 20% of a teacher's salary contingent on her performance.) Figure 9 shows the output and cost effects of raising the retention threshold so that ever larger shares of current teachers would be fired.²³ Productivity benefits scale roughly linearly. However, the costs increase nonlinearly, reaching about 50% of the total salary pool when the share of current teachers who would be

²²This is based on teacher salaries representing about one-third of total educational expenditures, an assumption that outcomes are linear in the log of class size, and Krueger's (1999) result that reducing class sizes from 22 to 15 raised scores by about 0.15 standard deviations. My labor supply model implies that hiring more teachers requires increasing all teachers' salaries, but this effect is negligible for class size reduction.

²³I consider only firing rates below 50%. When f^F exceeds 0.5, the retention threshold y^F rises above zero, and very many first year teachers – for whom the district's posterior mean equals 0.4 times the first-year performance – are fired immediately.

fired approaches 45%.

The dashed lines in Figure 9 present alternative simulations that assume lower labor supply elasticities ($\eta = \zeta = 1.5$) and risk aversion ($u(w) = \frac{1}{1-\rho} w^{1-\rho}$, with relative risk aversion parameter $\rho = 3$) on the part of teachers. This has relatively little effect on the productivity benefits of increased firing rates, but more dramatic effects on the costs: With these parameters – arguably more realistic than my baseline – the costs explode when f^F rises above about 30%. Importantly, costs become quite large even with much lower firing rates. Setting a retention threshold that would exclude one-fifth of current teachers would require increasing the teacher salary budget by more than 20%.

4.3 Sensitivity to alternative parameters & policies

Of course, all of the results presented above are dependent on the specific parameter values set out in Table 1. Table 3 presents estimates of the achievement effects and costs of the policies – at the original scale – under alternative parameter values. The first row repeats the estimates for the baseline parameters. Rows 2 and 3 expand the amount of private information that prospective teachers have about their own abilities. In row 2, a prospective teacher has the equivalent of one annual performance measure, while in row 3 she has the equivalent of two annual signals. More private information leads to larger achievement effects. It reduces costs under the firing policies – it allows teachers to better predict firing outcomes, thereby permitting those who would be fired to select out beforehand. But it raises costs slightly under the bonus contract, under which more bonuses would need to be paid out.

Row 4 shows estimates for a less noisy performance measure, with reliabil-

ity 0.6 in place of the 0.4 used for the earlier results. This makes the bonus contract much more effective, but has little effect on the firing contract.

Rows 5 through 10 show the effects of varying the labor supply elasticities. In general, both policies are more effective when labor supply is more elastic. The bonus policy is more sensitive to the exit than to the entry elasticity and becomes more expensive as the elasticities rise. Both relationships are of the opposite sign for the firing policy.

Row 11 shows an additional variant in which the entry elasticity is specified to be an increasing function of μ : $\eta = 3 + 2\frac{\mu}{\sigma_\tau}$. In a Roy model with $corr(\mu, \omega_1) > 0$, entry of higher- μ potential teachers is more elastic than that of those with lower μ . The function here is approximately what would obtain with a correlation around 0.5. It somewhat shrinks the costs of the policies, with little effect on the benefits.

Rows 12 and 13 show the impact of risk aversion on the results. In row 12, I assume that teachers are risk averse, with constant relative risk aversion parameter 3, over their annual incomes. One drawback of this is that it treats that idiosyncratic annual risk (such as is generated by noise in bonus receipt) as just as costly as is permanent risk like that generated by inaccurate firing policies. To partially address this, in Row 13 I modify equation (3) to allow teachers to be risk averse over the sum of their current income and their continuation values:

$$V_t(\theta_{t-1}; C) = \left(E \left[(w_t + \delta \max(\omega_{t+1}, V_{t+1}(\theta_t; C)))^{\frac{1}{1-\rho}} \middle| \theta_{t-1} \right] \right)^{1-\rho}. \quad (5)$$

This does not fully capture consumption smoothing, but it is a step in the right

direction. Risk aversion of this type leads to much smaller cost increases and better self-selection under the bonus policy, but makes firing a bit more expensive.

Row 14 shows an alternative bonus contract in which larger bonuses are given to fewer teachers. This attenuates the effect on student achievement. Row 15 shows estimates when firing is more costly to the worker – this makes the firing policy more effective but also much more expensive.

The appendix presents results for two additional contracts. One modifies the bonus policy to condition on the full performance history rather than on just the two most recent measures. This makes the policy dramatically more effective. The second is a tenure policy in which firing decisions are made only after a teacher's second year. This is less effective and more expensive than my firing policy, but the differences are small.

5 Misalignment of Performance Measure & Goal Distortion

I have assumed thus far that the performance measure is a noisy but otherwise accurate measure of teacher ability. But this overlooks two important sources of slippage between true and measured output. First, teachers' output is multidimensional – they should raise students' math and reading scores, but should also teach non-cognitive skills, other academic subjects (e.g., history, science, etc.), and non-academic topics like citizenship and art. Even an excellent performance measure is likely to capture the full range of outputs only imperfectly. Second, teachers facing strong incentives may be able to raise their measured performance without improving their overall productivity, by redirecting effort from unmeasured to measured dimensions (Glewwe et al., 2010) or simply by distorting the performance measure

directly, such as by cheating (Jacob and Levitt, 2003) or teaching to the test.

In this Section, I explore the implications of these issues under the firing contract. Results are necessarily extremely speculative, as little is known about either factor. Unfortunately, efforts to correlate measured performance with true productivity are severely hampered by the lack of an agreed-upon, comprehensive definition of productivity. But there is suggestive evidence that the correlation may not be very high. The Gates Foundation’s Measures of Effective Teaching (MET) project, for example, has found that teachers’ value added – net of measurement error – for students’ scores on traditional standardized tests is correlated only 0.37 to 0.54 with the teachers’ effects on student scores on more cognitively demanding, open response exams (Bill & Melinda Gates Foundation, 2010; Rothstein, 2011).²⁴ Correlations between value-added measures and teacher observations are generally even lower than this (Bill & Melinda Gates Foundation, 2012). It thus seems implausible that any feasible performance measure will be very highly correlated with a comprehensive understanding of a teacher’s true productivity.

I first augment the model developed above to incorporate this. I assume that each teacher performs two tasks and that her ability to perform the first, τ_{iA} , is imperfectly correlated with her ability to perform the second, τ_{iB} . I further assume that the two are jointly normal with identical variances and that the performance measure is based on only the first output dimension, $y_{it} = \tau_{iA} + \varepsilon_{it}$.

The first column of Table 4 presents baseline estimates when $\text{corr}(\tau_{iA}, \tau_{iB}) = 1$.²⁵ In the second column, I assume that $\text{corr}(\tau_{iA}, \tau_{iB}) = 0.4$, consistent with the

²⁴These are “disattenuated,” intended to estimate $\text{corr}(\tau_A, \tau_B)$ rather than $\text{corr}(y_A, y_B)$.

²⁵Reported results in Table 4, column 1 differ slightly from those seen earlier. The augmented model incorporating influence activities (discussed below and reported in columns 3-4) is more computationally intensive and as a result I use a less accurate numerical approximation. See the

MET evidence.²⁶ Because the firing policy selects (imperfectly) only on τ_{iA} and because $E[\tau_{iB} | \tau_{iA}] < \tau_{iA}$, the effect of the policy on the second dimension of teacher output is less than half as large as that on the measured dimension.

This assumes, as I have so far, that teachers' productivity is exogenous and unalterable. But it is natural to expect that teachers have some latitude to distribute their efforts across the different dimensions of output. If so, high-stakes incentives based on one of the dimensions will cause teachers to focus on that dimension, even if that comes at the exclusion of the other.

Essentially nothing is known about the quantitative magnitude of goal distortions and other influence activities in teaching.²⁷ Nevertheless, it seems important to understand whether such distortions can plausibly be important components of the response to high-stakes incentives, and even more so whether they undercut the intended effects. I thus adopt an extremely *ad hoc* model of the teacher's effort response. I assume that a teacher can each year choose an effort level E to be devoted to influencing the performance measure, producing an output measure $y_{it} = \tau_{iA} + E_{it} + \varepsilon_{it}$, but that the teacher must pay a cost $c(E) = kE^2$ to do so. I choose k so that raising measured performance by one standard deviation of the conditional distribution of τ_{iA} given τ_{iB} costs 20% of a first year teacher's annual

Appendix for details. I use the relatively inaccurate approximation in columns 1-2 of Table 4 as well, to facilitate comparisons across models.

²⁶I assume that μ is unidimensional and that $E[\tau_{iA} | \mu_i] = E[\tau_{iB} | \mu_i] = \mu_i$. If prospective teachers had any information about their relative effectiveness on the two dimensions, the effect on dimension-B output would be even smaller.

²⁷Carrell and West (2010) present suggestive evidence from a very different context that it could be important: They show that adjunct math instructors at the Air Force Academy produce better outcomes on the measure on which they are evaluated than do tenured faculty but that the adjuncts' students do worse in the long run. See also Glewwe et al. (2010), Campbell (1979), and Figlio and Loeb (2011).

salary.²⁸ This is extremely high – most forms of influence activities would be much less personally costly than this.

Teachers choose E to trade off the costs $c(E)$ and the benefits of distorting their measured performance, which depend on performance to date and on the teacher’s information about her own ability. Manipulation raises measured performance but makes it harder for the district to identify and fire the weakest teachers. As a consequence, the benefits of the firing policy, net of the manipulation, are attenuated. This is shown in column 3 of Table 4. The impact of the firing policy on average *measured* effectiveness is slightly lower than without manipulation. But when the effects of distortionary effort are excluded – which they should be if the influence activity takes an unproductive form such as cheating or teaching to the test – the policy’s impact shrinks by a third. And the effects on dimension-B effectiveness are even smaller, only 0.013 student-level standard deviations.

Even this may understate the degree to which distortionary effort can undercut the firing policy. Many forms of manipulation – e.g., narrowing of the curriculum or diversion of class time to test preparation – take the form of transferring output from unmeasured dimensions to those covered by the test. In column 4, I present results when effort E reduces dimension-B output one-for-one with its positive effect on measured dimension-A output. In this case, the impact of the firing policy on dimension-B output is almost totally eliminated: The negative consequences of teaching to the test offset nearly all of the improved selection on τ_{iB} seen in column 3. Despite this, the district must continue to pay a substantial cost in higher salaries to compensate teachers for the costs they bear from their manipu-

²⁸With $\text{corr}(\tau_{iA}, \tau_{iB}) = 0.4$ and $SD(\tau_{iA}) = SD(\tau_{iB}) = 0.15$, the standard deviation of τ_{iA} given τ_{iB} is 0.137, and $k = 10.6w^0$.

lation effort and from the firing that persists.

It must be emphasized that the effort model here is not based on any specific evidence of the cost or quantity of manipulation of the performance measure in response to high-stakes incentives. But the importance of manipulative activity in response to high-stakes incentives in education is well established and it seems quite plausible that distortions of the measurement process could be even worse than is assumed here. Understanding their form and quantitative magnitude is evidently extremely important to predicting even the qualitative impact of teacher quality policies.

6 Discussion

The simulations presented here suggest that the effects of many proposed teacher quality policies will depend importantly on their interaction with the teacher labor market. So long as prospective teachers are uncertain about their own abilities or labor supply to teaching is less than perfectly elastic, both performance-based compensation and performance-based retention policies require substantial increases in total teacher compensation in order to produce meaningful changes in productivity.

Assuming that the necessary funding is available and that teachers are unable to game the performance measurement process, both classes of policies appear to be cost effective at modest scales relative to “traditional” uses of additional funds. Indeed, recognition of the labor market effects can make non-retention policies even more effective than when these effects are ignored, as the accompanying salary increases help to attract and retain high ability teachers. My results also point to the importance of policy design, as cost-effectiveness varies importantly with the

specifics of the teacher contract.

There are several important caveats to my results, however. First, and most importantly, they rely on a best case view of the potential for teacher performance assessment. As Section 5 shows, effects on unmeasured dimensions of teacher productivity are likely to be much weaker than those on measured performance. Moreover, even these effects depend crucially on the assumption that performance measures are noisy but incorruptible. In the real world, every performance measure is susceptible to “influence activities” that raise the measure out of proportion to changes in true performance. If teachers can improve their measured performance by arranging to have the right students, by reducing the attention paid to non-tested topics and subjects, by teaching to the test, or by outright cheating, then the improvements in true learning that would obtain under high-stakes accountability policies are dramatically attenuated. Effects on unmeasured dimensions of productivity could easily be negative. Two high priority topics for future research must be the degree to which available performance measures are correlated with other dimensions of teacher output and the extent to which the measures become corrupted when the stakes are raised (Rothstein, 2011).²⁹

Even when the possibility of systematic divergence between measured and true effectiveness is ruled out by assumption, the impacts of alternative teacher contracts on student achievement are modest. Neither of the alternative contracts considered here would raise average productivity by more than one-third of a standard deviation, even under extremely optimistic parameters. These kinds of benefits

²⁹Chetty et al. (2011) relate teachers’ value-added to students’ later earnings. But their analysis can only show that the correlation is positive; they do not estimate the magnitude. And their estimates derive from a low-stakes setting.

would be most welcome, but would not represent fundamental changes in our education system. Larger benefits are possible if the scope of the policies is increased, but only with prohibitive increases in the teacher salary bill.

Finally, the calibration results of course depend importantly on my choice of parameter values. In particular, if labor supply to the teaching profession is less elastic than I have assumed — based on district-level studies that should be expected to provide upper-bounds to the occupation-level elasticity — then each policy becomes much less effective.

There of course a number of important aspects of the teaching profession that are omitted from my stylized model. I have already discussed the potential for influence activities aimed at gaming the performance measure. Another omitted consideration is the role of pre-service training as a component of the teaching career. This can be seen as a fixed cost of entering the profession. Performance-based retention policies would be much more expensive in the presence of large fixed costs. Thus, my analyses that abstract from such costs almost certainly overstate the benefits of these policies. They can perhaps be seen as validation for the claim sometimes made by advocates of performance-based retention policies (e.g., Staiger and Rockoff, 2010; Gordon et al., 2006) that the cost to prospective teachers of increased riskiness can be offset by reducing certification requirements, though this claim rests importantly on the hypothesis that these requirements do nothing either to screen out low ability potential teachers.

A related issue is that of variation in hours of work over the career. Insofar as early career teachers invest heavily in preparing lesson plans that they will reuse later in their careers, the effective hourly wage in teaching is quite low at the begin-

ning of the career and higher at the end. This age profile is further accentuated by the backloading of teacher compensation through generous pensions and often quite steep salary schedules. Like certification requirements, backloaded compensation raises the cost to a teacher of early career displacement, as it means that she will never be able to collect the high effective hourly wages given to experienced teachers, and thus makes the profession much less attractive if firing is a real possibility. It is less clear how this sort of fixed cost could be reduced.

These caveats aside, the analysis here demonstrates that clear thinking about the potential impact of teacher quality policy requires a careful, accurate model of the roles of imperfect information and teacher labor supply decisions. More research is needed on these factors, and on their impact on the optimal design of the teacher contract. For now, though, it seems safe to conclude that plausible policies aimed at changing the ability distribution of the teacher workforce through improved selection are unlikely to have dramatic impacts on student achievement absent a performance measurement system that is immune to manipulation and that is accompanied by substantial increases in the resources devoted to teacher pay.

References

Adda, Jerome, Christian Dustmann, Costas Meghir, and Jean-Marc Robin,

“Career Progression and Formal versus On-the-Job Training,” April 6 2011. Unpublished manuscript.

Baker, George P., “Incentive contracts and performance measurement,” *Journal of Political Economy*, 1992, 100 (3), 598–614.

— , “Distortion and risk in optimal incentive contracts,” *The Journal of Human Resources*, 2002, 37 (4), 728.

Ballou, Dale, “Do public schools hire the best applicants?,” *The Quarterly Journal of Economics*, 1996, 111 (1), pp. 97–133.

Bill & Melinda Gates Foundation, “Learning about teaching: Initial findings from the Measures of Effective Teaching project,” Research Report, Bill & Melinda Gates Foundation, Seattle, Washington 2010.

— , “Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains,” Research Paper, Bill & Melinda Gates Foundation, Seattle, Washington 2012.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff, “Teacher layoffs: An empirical illustration of seniority versus measures of effectiveness,” *Education Finance and Policy*, 2011, 6 (3), 439–454.

Campbell, Donald T., “Assessing the impact of planned social change,” *Evaluation and Program Planning*, 1979, 2, 67–90.

Carrell, Scott E. and James E. West, “Does professor quality matter? Evidence from random assignment of students to professors,” *Journal of Political Economy*, 2010, 118 (3), 409–432.

Chandler, Michael Alison, “Schools Pinched in Hiring,” *The Washington Post*, June 24 2007.

Chetty, Raj, John Friedman, and Jonah E. Rockoff, “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood,” Working Paper No. 17699, National Bureau of Economic Research, December 2011.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, “How and Why Do Teacher Credentials Matter for Student Achievement?,” Working Paper 12828, National Bureau of Economic Research, January 2007.

Corcoran, Sean, “Can Teachers Be Evaluated by Their Students’ Test Scores?,” Education Policy for Action Series, Annenberg Institute for School Reform, September 2010.

Corcoran, Sean P., William N. Evans, and Robert M. Schwab, “Changing labor-market opportunities for women and the quality of teachers, 1957-2000,” *The American Economic Review*, 2004, 94 (2), pp. 230–235.

Dolton, Peter and Wilbert van der Klaauw, “The turnover of teachers: A competing risks explanation,” *Review of Economics and Statistics*, 1999, 81 (3), 543–550.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan, “Incentives work: Getting teachers to come to school,” *American Economic Review*, June 2012, 102 (4), 1241–1278.

Figlio, David and Susanna Loeb, “School Accountability,” in Eric Hanushek, Stephen Machin, and Ludger Woessman, eds., *Handbook of the Economics of Education*, Vol. 3, Elsevier, 2011, chapter 8, pp. 383–421.

Flyer, Fredrick and Sherwin Rosen, “The new economics of teachers and education,” *Journal of Labor Economics*, 1997, 15 (1), pp. S104–S139.

Fryer, Roland G., “Teacher Incentives and Student Achievement: Evidence from New York City Public Schools,” Working paper 16850, National Bureau of Economic Research, 2011.

Fryer, Roland G, Steven D. Levitt, John List, and Sally Sadoff, “Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment,” Working Paper 18237, National Bureau of Economic Research, July 2012.

Gates, Bill, “How teacher development could revolutionize our schools,” *The Washington Post*, February 28 2011.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, “Teacher incentives,” *American Economic Journal: Applied Economics*, July 2010, 2 (3), 205–227.

Goldhaber, Dan, Betheny Gross, and Daniel Player, “Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best?,” *Journal of Policy Analysis and Management*, 2011, 30 (1), 57–87.

Goodman, Sarena F. and Lesley J. Turner, “Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program,” *Journal of Labor Economics*, Forthcoming.

Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger, “Identifying effective teachers using performance on the job,” Discussion Paper 2006-01, The Hamilton Project, April 2006.

- Hanushek, Eric A. and Steven G. Rivkin**, “Teacher quality,” in Eric Hanushek and Finis Welch, eds., *Handbook of the economics of education*, Vol. 2, Elsevier, 2006, pp. 1051–1078.
- Hess, Rick**, “Missing the POINT: Tomorrow’s big merit pay study will tell us...nothing,” *Education Week: Rick Hess Straight Up [blog]*, September 2010.
- Hiatt, Fred**, “How Bill Gates would repair our schools,” *The Washington Post*, March 30 2009.
- Holmstrom, Bengt and Paul Milgrom**, “Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design,” *Journal of Law, Economics and Organization*, January 1991, 7 (Spring), 24–52.
- Jacob, Brian A. and Steven Levitt**, “Rotten apples: An investigation of the prevalence and predictors of teacher cheating,” *Quarterly Journal of Economics*, 2003, 118 (3), 843–847.
- Jacobson, Stephen L.**, “Monetary incentives and the reform of teacher compensation: A persistent organizational dilemma,” *International Journal of Educational Reform*, January 1995, 4 (1), 29–35.
- Keane, Michael P. and Kenneth I. Wolpin**, “The career decisions of young men,” *Journal of Political Economy*, June 1997, 105 (3), 473–522.
- Klein, Joel, Michelle Rhee, Peter C. Gorman, Ron Huberman, Carol R. Johnson, Andres A. Alonso, Tom Boasberg, Arlene C. Ackerman, William R. Hite Jr., Jean-Claude Brizard, Jose M. Torres, J. Wm. Covington, Terry B. Grier,**

- Paul Vallas, Eugene White, and LaVonne Sheffield**, “How to Fix Our Schools: A Manifesto by Joel Klein, Michelle Rhee and Other Education Leaders,” *The Washington Post*, October 10 2010.
- Krieg, John M.**, “Teacher quality and attrition,” *Economics of Education Review*, 2006, 25 (1), 13 – 27.
- Krueger, Alan B.**, “Experimental estimates of education production functions,” *Quarterly Journal of Economics*, 1999, 114 (2), 497–532.
- Lavy, Victor**, “Evaluating the effect of teachers’ group performance incentives on pupil achievement,” *Journal of Political Economy*, December 2002, 110 (6), 1286–1317.
- Lazear, Edward P.**, “Performance pay and productivity,” *American Economic Review*, 2000, 90 (5), 1346–1361.
- , “Teacher incentives,” *Swedish Economic Policy Review*, 2003, 10 (2), 179–214.
- Manning, Alan**, *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton, NJ: Princeton University Press, 2005.
- Mathematica Policy Research**, “Information about the national evaluation of the Teacher Incentive Fund (TIF),” undated. Accessed from <http://mathematica-mpr.org/publications/pdfs/TIFEvaluationInfo.pdf> on August 23, 2011.

Murnane, Richard J. and David K. Cohen, “Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive,” *Harvard Educational Review*, 1986, 56 (1), 1–17.

— **and Randall J. Olsen**, “The effect of salaries and opportunity costs on duration in teaching: Evidence from Michigan,” *The Review of Economics and Statistics*, May 1989, 71 (2), 347–352.

Ransom, Michael R. and David P. Sims, “Estimating the firm’s labor supply curve in a ‘new monopsony’ framework: Schoolteachers in Missouri,” *Journal of Labor Economics*, April 2010, 28 (2), 331–355.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, “Teachers, schools, and academic achievement,” *Econometrica*, 2005, 73 (2), 417–458.

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*, 2011, 6 (1), 43–74.

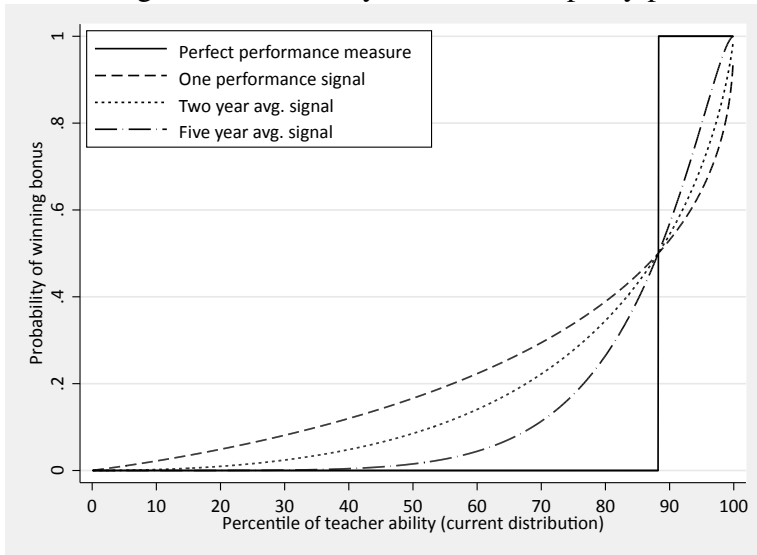
Rothstein, Jesse, “Student sorting and bias in value-added estimation: Selection on observables and unobservables,” *Education Finance and Policy*, 2009, 4 (4), 537–571.

—, “Teacher quality in educational production: Tracking, decay, and student achievement,” *Quarterly Journal of Economics*, 2010, 125 (1), 175–214.

—, “Review of ‘Learning about teaching: Initial findings from the Measures of Effective Teaching project’,” National Education Policy Center, Boulder, CO, 2011.

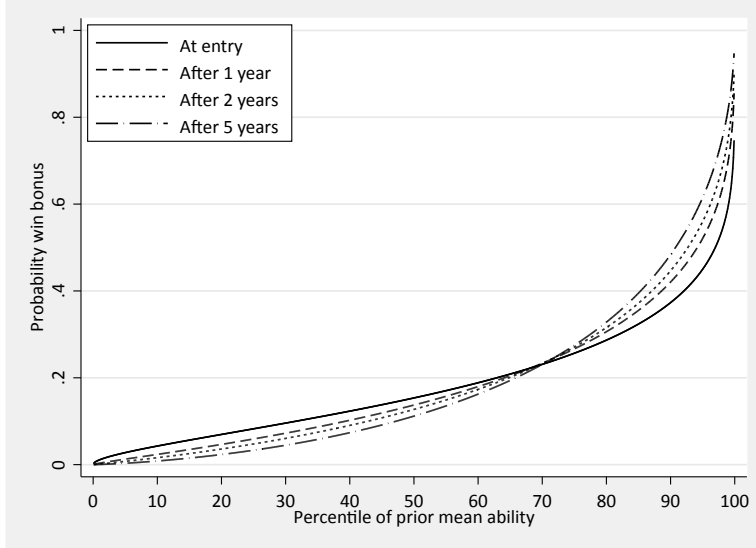
- Sass, Tim R.**, “The stability of value-added measures of teacher quality and implications for teacher compensation policy,” Policy Brief 4, National Center for Analysis of Longitudinal Data in Education Research, November 2008.
- Springer, Matthew G., Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher**, “Teacher pay for performance: Experimental evidence from the Project On Incentives In Teaching,” National Center for Performance Incentives at Vanderbilt University, 2010.
- Staiger, Douglas O. and Johah E. Rockoff**, “Searching for effective teachers with imperfect information,” *Journal of Economic Perspectives*, 2010, 24 (3), 97–118.
- Taylor, Eric S. and John H. Tyler**, “The Effect of Evaluation on Teacher Performance,” *American Economic Review*, Forthcoming.
- Tincani, Michela**, “Teacher Labor Markets, School Vouchers and Student Cognitive Achievement: Evidence from Chile,” December 2011. Unpublished manuscript.
- von Wachter, Till, Jae Song, and Joyce Manchester**, “Long-term earnings losses due to mass layoffs during the 1982 recession: An analysis using U.S. administrative data from 1974 to 2004,” April 2009. Unpublished manuscript.
- Wiswall, Matthew**, “Licensing and Occupational Sorting in the Market for Teachers,” 2007. Unpublished manuscript.

Figure 1: Probability of bonus receipt, by percentile of true ability



Notes: Graph shows the probability that a teacher with ability τ (scaled as $p = 100\Phi(\tau/\sigma_\tau)$) would have average performance over 1, 2, or 5 years above $y^B = 0.178$.

Figure 2: Probability of bonus receipt, by prior percentile and years of experience



Notes: Horizontal axis is scaled as percentile of the t -specific distribution of $\hat{\tau}_{it}$, for $t = 0, 1, 2, 10$, under the baseline contract. Vertical axis shows the probability that a teacher with each prior mean would have average measured performance over 2 years above $y^B = 0.178$.

Figure 3: Average perceived probability of bonus receipt, by percentile of true ability and years of experience

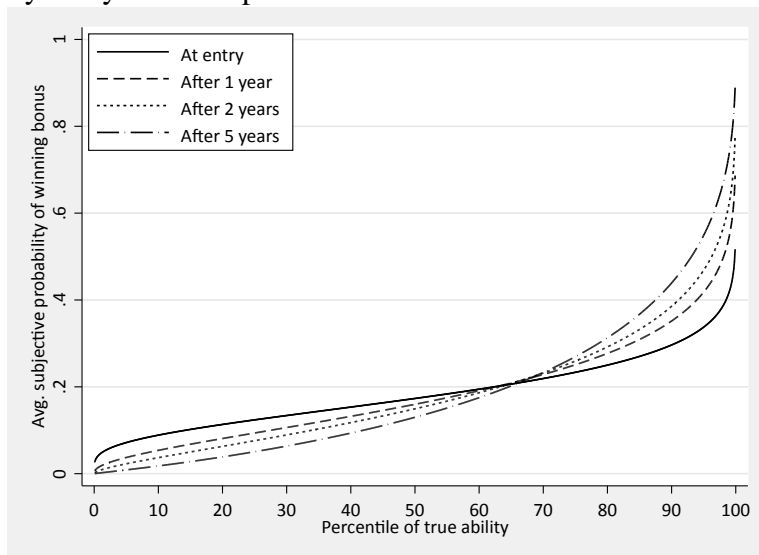
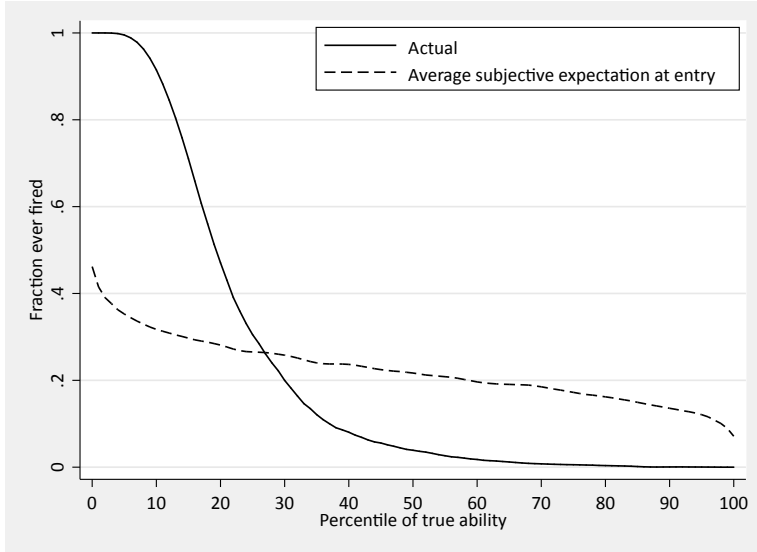
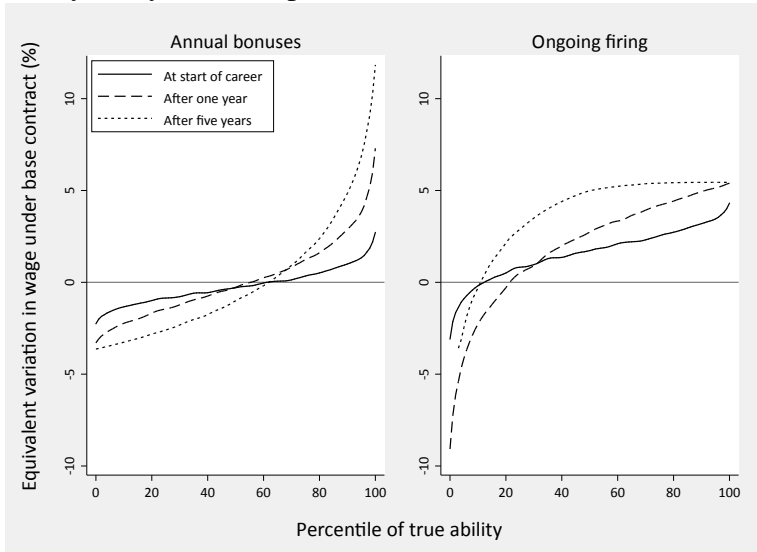


Figure 4: Probability of ever being fired (assuming no quits) and average subjective expectation of firing probability at start of career, by percentile of true ability



Notes: Figure shows fraction of teachers who would be fired before the end of a 30-year teaching career, as well as the average subjective probability of such event as of the beginning of the teaching career.

Figure 5: Effects of alternative contracts on average value, by percentile of true ability and years of experience



Notes: Values expressed as equivalent variation in base pay (w^0) relative to the base contract. Values are averaged only across teachers not yet fired.

Figure 6: Effects of alternative contracts on number of new entrants to teaching, by percentile of true ability

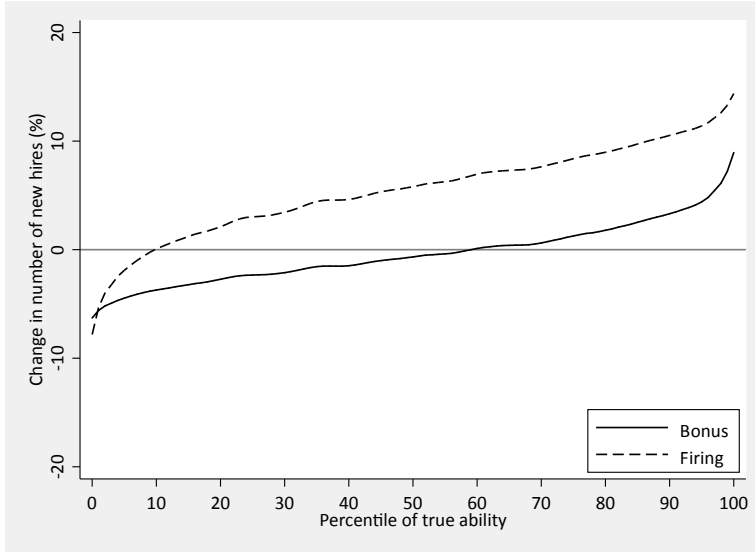


Figure 7: Effects of alternative contracts on average teaching career lengths, by percentile of true ability

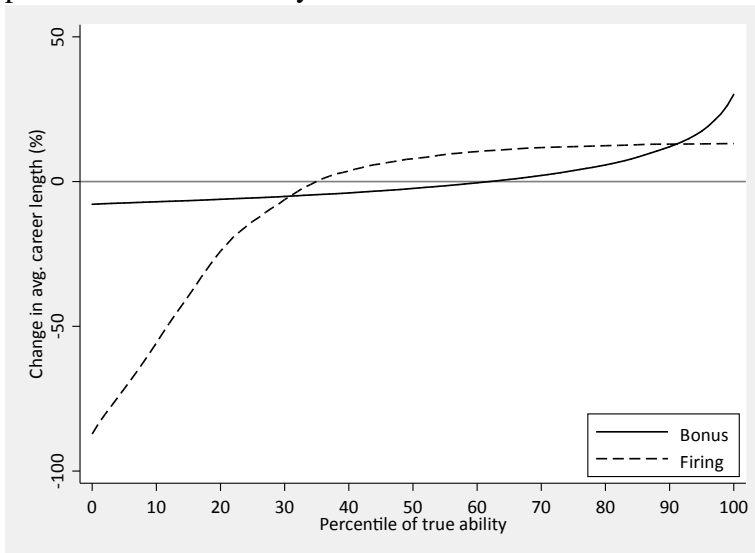


Figure 8: Effects of alternative contracts on total number of teachers, by percentile of true ability

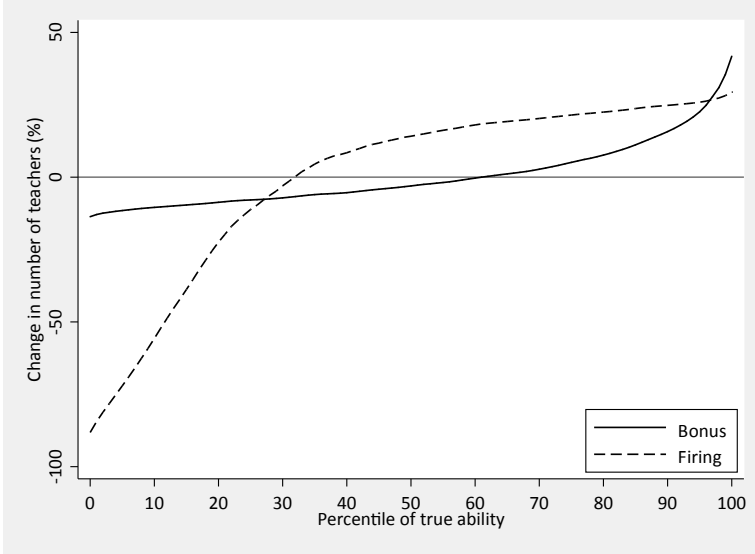
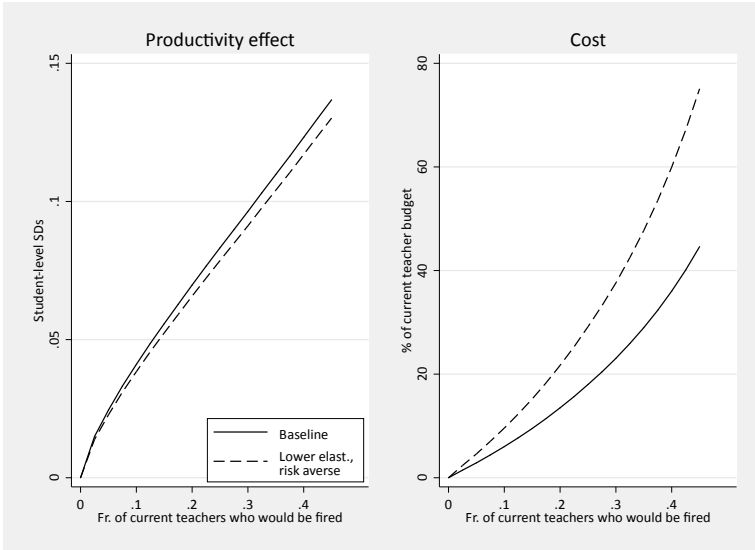


Figure 9: Effect of tenure and firing contracts on average output and total costs, by share not tenured or fired



Notes: Dashed line corresponds to $\eta = \zeta = 1.5$ and $u(w) = \frac{1}{1-\rho}w^{1-\rho}$, with $\rho = 3$.

Table 1: Key parameters and base values

Category / parameter	Description	Baseline value
Effectiveness		
σ_τ	SD of teacher effectiveness	0.15
$r(t)$	Experience effect on productivity	$\begin{cases} -0.07 & \text{if } t = 0 \\ -0.04 & \text{if } t = 1 \\ -0.02 & \text{if } t = 2 \\ 0 & \text{if } t > 2 \end{cases}$
Measurement		
σ_ε	SD of noise in annual performance measure	0.18
Teacher preferences & information		
h	Reliability of pvt. info. as measure of ability	0.25
δ	Discount rate (real)	0.97
$u(\cdot)$	von Neumann-Morgenstern utility function	$u(w) = w$
η	Elasticity of entry with respect to w_0	3
ζ	Negative of elasticity of exit hazard w.r.t. w_0	3
λ_0	Annual exit hazard under base contract	0.08
T	Maximum length of teaching career (years)	30
κ	Effect of being fired	10%
Base contract		
$g(x)$	Real return to experience	$0.015 * x$
Bonus contract		
b	Bonus size (as share of base pay)	20%
f^B	Fr. of current teachers who would receive bonus	25%
α^B	Base pay as share of pay under baseline contract	96.4%
Firing contract		
f^F	Fr. of current teachers who would be fired	10%
α^F	Base pay as share of pay under baseline contract	105.4%

Table 2: Impact of bonus and firing contracts on teacher effectiveness and total costs

	Baseline	Bonuses		Firing	
		Level	Change from baseline	Level	Change from baseline
	(1)	(2)	(3)	(4)	(5)
Teacher ability (τ)					
Mean	0.000	0.015	+0.015	0.040	+0.041
SD	[0.150]	[0.153]	+0.003	[0.130]	-0.020
Teacher experience					
Pct. 1st year	8.0%	8.0%	-0.0 p.p.	8.1%	+0.1 p.p.
Pct. 1st 3 years	30.9%	30.8%	-0.1 p.p.	31.0%	+0.1 p.p.
Mean	8.82	8.87	+0.045	9.11	+0.284
Teacher effect ($\tau+r(t)$)					
Mean	-0.011	0.004	+0.015	0.029	+0.040
SD	[0.151]	[0.155]	+0.004	[0.134]	-0.017
Salaries (expressed as multiple of baseline starting salary)					
Base starting salary	1.000	0.964	-3.6%	1.054	+5.4%
Average total pay	1.148	1.168	+1.8%	1.216	+5.9%

Table 3: Sensitivity of results to alternative parameters

Row	Bonuses		Firing	
	Effect on output (student SDs)	Effect on salary bill (%)	Effect on output (student SDs)	Effect on salary bill (%)
	(1)	(2)	(3)	(4)
1 Baseline	+0.015	+1.8%	+0.040	+5.9%
More private information				
2 $h=0.4$	+0.018	+1.8%	+0.042	+5.8%
3 $h=0.57$	+0.022	+1.9%	+0.044	+5.5%
Less noisy performance measure				
4 $\sigma_\varepsilon=0.12$	+0.021	+2.1%	+0.041	+5.7%
Varying the supply elasticities				
5 $\eta=6, \zeta=3$	+0.019	+1.9%	+0.044	+4.9%
6 $\eta=3, \zeta=6$	+0.026	+2.1%	+0.042	+4.7%
7 $\eta=6, \zeta=6$	+0.031	+2.3%	+0.045	+4.3%
8 $\eta=1.5, \zeta=3$	+0.013	+1.7%	+0.039	+7.1%
9 $\eta=3, \zeta=1.5$	+0.009	+1.6%	+0.039	+7.2%
10 $\eta=1.5, \zeta=1.5$	+0.007	+1.5%	+0.038	+9.4%
11 η rising with $\mu, \zeta=3$	+0.015	+1.3%	+0.041	+5.3%
With risk aversion				
12 Cash-in-hand	+0.012	+2.5%	+0.040	+6.0%
13 Over PDV of lifetime income	+0.014	+1.9%	+0.041	+6.2%
Other variations				
14 50% bonus to 10% of teachers	+0.011	+1.6%		
15 Firing reduces wages by 20%			+0.045	+7.9%

Table 4: Firing policy when output is multidimensional and performance measure is corruptible

	Baseline	Multi-dimensional output		
		No influence	With costly influence	With costly & counter-productive influence
	(1)	(2)	(3)	(4)
Key parameters				
corr(τ_1, τ_2)	1	0.4	0.4	0.4
Influence possible?	n	n	y	y
Influence is counterproductive	n	n	n	y
Impact of firing policy on:				
Effectiveness on dimension A				
As measured	+0.042	+0.042	+0.037	+0.037
Net of influence	n/a	n/a	+0.030	+0.030
Effectiveness on dimension B	+0.042	+0.020	+0.013	+0.005
Average total pay	+6.1%	+6.1%	+3.9%	+3.9%

A Appendices

A.1 Search model

Each teacher draws a single outside job offer each year. If she accepts the offer, she exits teaching forever. The outside offer arrives after the teacher learns her previous year's performance (and is paid on that basis).

Outside offers are indexed by the continuation value that they provide, ω . I assume that the outside offer ω_t , $t > 1$ has a censored Pareto distribution:

$$F_t(\omega_t) = \begin{cases} 0 & \text{if } \omega_t \leq V_t^0 \lambda_0^{1/\zeta'_t} \\ 1 - \lambda_0 \left(\frac{V_t^0}{\omega_t}\right)^{\zeta'_t} & \text{if } V_t^0 \lambda_0^{1/\zeta'_t} < \omega_t < HV_t^0 \\ 1 & \text{if } HV_t^0 \leq \omega_t. \end{cases} \quad (\text{A.1})$$

Here, V_t^0 is the value obtained under the baseline, single salary contract (which is constant across teachers), λ_0 is the annual exit hazard under this contract, and H is the maximum outside wage, expressed as a fraction of the inside continuation value.³⁰ Importantly, the distribution of ω_t is independent of the teacher's ability as a teacher, τ_i . Thus, as the teacher learns about τ_i she does not simultaneously learn about her future outside options.

Under the outside distribution (A.1), the probability that a teacher who would obtain continuation value $V_t \in [V_t^0 \lambda_0^{1/\zeta'_t}, HV_t^0]$ in teaching will instead exit is $\lambda_t(V_t) = \Pr\{\omega_t > V_t\} = \lambda_0 (V_t^0/V_t)^{\zeta'_t}$, with $\partial \ln \lambda_t(V_t)/\partial \ln V_t = -\zeta'_t$. The model in the main text was developed in terms of the negative of the elasticity of the exit hazard with respect to the inside wage under the baseline contract, $\zeta \equiv -\partial \ln \lambda_t/\partial \ln w^0 = -\partial \ln \lambda_t/\partial \ln V_t * \partial \ln V_t/\partial \ln w^0 = \zeta'_t * \partial \ln V_t/\partial \ln w^0$. The latter fraction varies with t . I thus solve recursively for this elasticity – which depends on ζ'_s , $s > t$, but not on ζ'_t itself – and use it to define the elasticity parameter in (A.1) as $\zeta'_t \equiv \zeta * (\partial \ln V_t/\partial \ln w^0)^{-1}$.

The distribution of the initial non-teaching offer, ω_1 , is similar to that of offers later in the career, though here the shape parameter is computed as $\zeta'_1 \equiv \eta (\partial \ln V_0/\partial \ln w^0)^{-1}$.

³⁰The use of a censored distribution ensures that V_t is finite for any ζ'_t . It has no effect on the results so long as the censoring point is high enough that offers at that point are always accepted. I set $H = 2$, satisfying this criterion.

A.2 Solving the model

Equation (3) does not have a closed-form solution, but for any specified contract it can be solved recursively. Under the learning model developed above, the distribution of period- t performance measure given θ_{t-1} is

$$y_t | \theta_{t-1} \sim \mathbb{N} \left(\hat{\tau}_{t-1}, \frac{1}{\frac{1}{(1-h)\sigma_\tau^2} + \frac{t-1}{\sigma_\varepsilon^2}} + \sigma_\varepsilon^2 \right). \quad (\text{A.2})$$

This is a univariate distribution that can easily be computed for any specified value of $\hat{\tau}_{t-1}$. Given $\hat{\tau}_{t-1}$ and y_t , computation of $\hat{\tau}_t$ is trivial.

The recursive solution thus has three steps. First, I compute $w_T^C(y_1, \dots, y_T)$, the final period wage under contract C as a function of the performance signals to date. Second, I compute the value of remaining in teaching in period T , $V_T(\theta_{T-1}; C)$, as a function of θ_{T-1} , by integrating w_T^C over the conditional distribution of y_T given by (A.2). Third, for each $t < T$, given estimates of $V_{t+1}(\theta_t; C)$ as a function of θ_t , I compute $w_t^C(y_1, \dots, y_t)$ for each possible y_t , then integrate over the distribution of y_t (and therefore of θ_t) given θ_{t-1} to obtain $V_t(\theta_{t-1}; C)$.

The state space θ_t is of dimension $t + 1$, creating a dimensionality problem for careers of reasonable length. Note, however, that each of the contracts considered above reduces the state space for computation of w_t^C from the t -dimensional distribution $\{y_1, \dots, y_t\}$ to a one- or two-dimensional distribution: $\{y_{t-1}, y_t\}$ for the performance pay contract and $\{\bar{y}_t\}$ for the firing contract. Meanwhile, the teacher's assessment of her own ability at the end of period $t - 1$ can be summarized either by the single variable $\hat{\tau}_{t,t-1}$ or by the pair $\{\mu, \bar{y}_{t-1}\}$. I can thus focus on state spaces of only two dimensions, $\tilde{\theta}_{t-1} = \{\hat{\tau}_{t-1}, y_{t-1}\}$ for the bonus contract or $\check{\theta}_{t-1} = \{\mu, \bar{y}_{t-1}\}$ for the firing contract. I approximate the joint distributions of these two-dimensional state variables and y_t with grids of 149^3 points spaced to have equal probability mass.³¹

Having computed $V_t(\theta_{t-1}, C)$ for each t , θ_{t-1} , and C , I simulate the impact of policies by drawing potential teachers from the $\{\mu, \tau\}$ distribution, then drawing performance measures $\{y_1, \dots, y_T\}$ for each. For each career, I compute θ_{t-1} and V_t at each year t , and use these to compute the effects of contract C on the probability of entering the profession and, conditional on entering, on surviving to year t . Note that I need not model the distribution of $\{\mu, \tau\}$ in the *population* of potential teachers – under my constant elasticity assumptions, changes in the returns to

³¹In the model of influence activities in Section 5, \bar{E}_{t-1} is an additional state variable, and moreover the optimal choice of E_t must be solved for numerically. I use 49^3 points for the ability parameters $\{\mu, \bar{y}_{t-1} - \bar{E}_{t-1}, y_t - E_t\}$ and 24 points each for \bar{E}_{t-1} and E_t .

teaching induce proportional changes in the amount of labor supplied to teaching by each type that do not depend on the number of people of that type in the population.

A.3 Market clearing

Alternative contracts may yield greater or lesser entry or persistence in aggregate. For example, adding performance bonuses without reducing base pay will yield more entry from high- μ teachers and greater persistence of high- \hat{v}_t teachers, without offsetting reductions from teachers with low μ or \hat{v}_t . Under each alternative contract, I compute the steady-state size of the teacher workforce, assuming that the contract has been in place for at least T years and that the same number of entering teachers have been hired in each year. I assume that the education system will require the same number of teachers under the alternative contracts as are required under the baseline contract; where my computation yields a larger or smaller workforce than in baseline, I assume that the base salary is adjusted upward or downward to yield the appropriate number of teachers. The α^B and α^F parameters in Table 1 are the adjustments required given the other parameters listed there; these are found via a numerical search algorithm.

A.4 Additional contracts

The bonus contract discussed in the main text closely resembles many that have been proposed and sometimes even implemented. The firing contract does not – most selective retention policies that have been proposed would condition retention decisions on just a short performance history, either at the time of an early tenure decision or when layoffs are necessitated by budget shortfalls. I nevertheless present the firing policy because it makes more intelligent use of the available information in the context of my model. But this makes for an unfair comparison between compensation and retention policies, as one is closer to optimal than the other.

Here, I present results for two additional contracts that help to clarify the comparison. The first is an up-or-out tenure policy, with a single retention decision made after year 2; the second is a pay-for-performance policy that conditions annual pay on the district’s full information about a teacher’s type.

In the tenure contract, any teacher for whom $(y_{i1} + y_{i2})$ exceeds a threshold y^T is given tenure and cannot be fired thereafter; teachers who are not tenured are fired. y^T is set so that a fraction f^T of current new teachers would receive tenure.

Base salaries are increased to attract enough teachers, and pay is not sensitive to performance: $w_{it}^T = \alpha^T w_{it}^0$, with $\alpha^T > 1$. This contract resembles the ones considered by Staiger and Rockoff (2010) and actual policy in New York and elsewhere. Like the bonus contract, it is prone to frequent errors, as two years of performance is not enough to accurately identify ability.

In the performance pay contract, pay in year t depends linearly on the teacher's average demonstrated performance to date, $\bar{y}_{i,t-1}$, with a slope that grows with t in proportion to the increasing reliability of $\bar{y}_{i,t-1}$ as a signal of τ . Specifically, compensation in year $t > 1$ for a teacher with average performance through $t - 1$ periods of $\bar{y}_{i,t-1}$ is

$$w_{it}^{PP} = \alpha^{PP} w_{it}^0 \left(1 + \beta^{PP} \frac{\sigma_\tau^2}{(t-1)^{-1} \sigma_\varepsilon^2 + \sigma_\tau^2} \bar{y}_{i,t-1} \right),$$

with no differentiation in the first year ($w_{i1}^{PP} = \alpha^{PP} w_{i1}^0$). The performance coefficient β^{PP} multiplies the same district posterior mean that is used as the basis of the firing policy in the main text. Although I am not aware of a district that has implemented a policy of this form, it bears some resemblance to the permanent raises awarded recently to high-performing teachers in Washington DC. The Washington raises cannot be taken away if a teacher does not continue to perform well, but under my performance pay contract they can; if the teacher consistently performs at the same high level her pay will continue to rise as the district becomes more confident in its assessment of her.

Table A.1 presents results for the original two contracts and for the two new contracts. It is clear that both compensation and retention policies benefit from making better use of the available information rather than from limiting attention to two years of performance. But the difference is much larger for the compensation-based than for the retention-based policies. The pay-for-performance policy clearly dominates even a carefully designed firing policy, with somewhat larger benefits and half the costs.

Table A.2 presents an analysis of the pay-for-performance contract with multidimensional output and the potential for manipulation. This suggests caution about the previous result: This policy is even more damaged by manipulation than is the firing policy. Under a firing contract, low ability teachers have the most incentive to manipulate the performance measure; many of them wind up being fired anyway. Under the pay-for-performance contract, it is the highest ability teachers, who plan to stay the longest, who are most likely to engage in manipulation, and they do much more of it. This does not damage the selection effects of the contract – note that the effects of the contract on dimension-B effectiveness are similar in

columns 2 and 3, where they were quite different for the firing contract in Table 4. But if the manipulation itself is costly to dimension-B output, as in column 4, this can more than offset the benefits of the policy coming from selection, reducing total productivity below what it is under the baseline contract.

Figure A.1: Empirical one-year attrition hazards from the 1999/00 Schools and Staffing Survey/Teacher Follow-Up Survey

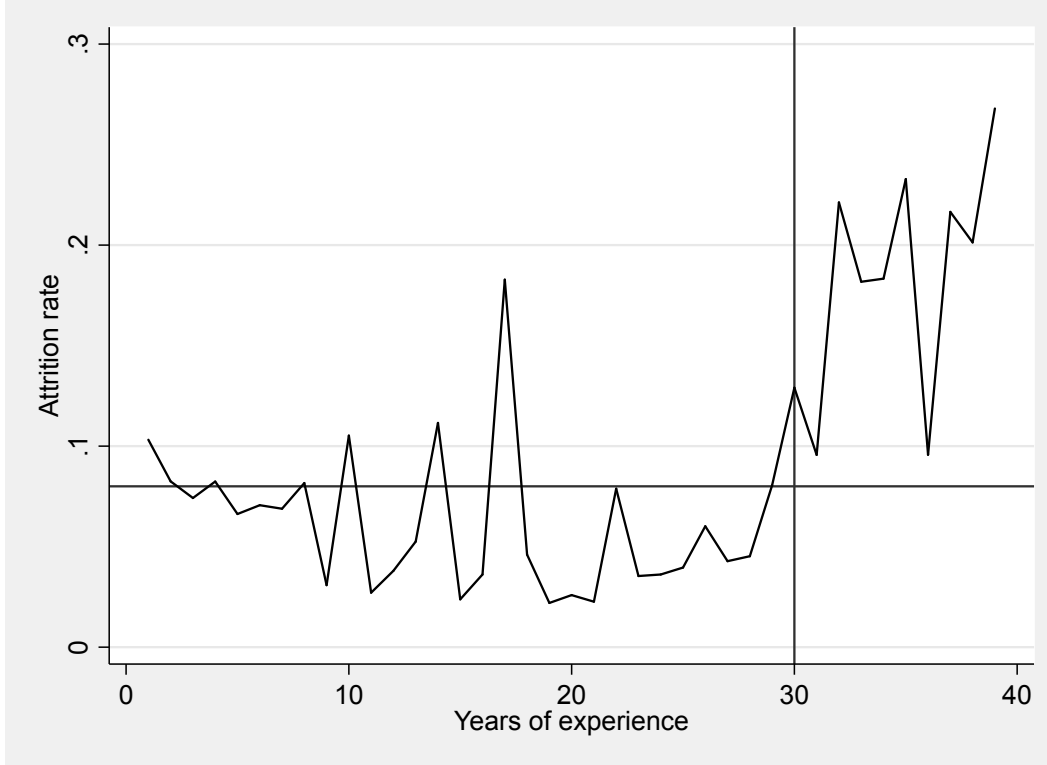


Table A.1: Impact of bonus, firing, performance pay, and tenure contracts on teacher effectiveness and total costs

	Baseline	Bonuses	Firing	Performance pay	Tenure
	(1)	(2)	(3)	(4)	(5)
<i>Outcomes under alternative contracts</i>					
Teacher ability (τ)					
Mean	0.000	0.015	0.040	0.046	0.037
SD	[0.150]	[0.153]	[0.130]	[0.150]	[0.135]
Teacher experience					
Pct. 1st year	8.0%	8.0%	8.1%	7.9%	8.5%
Pct. 1st 3 years	30.9%	30.8%	31.0%	30.5%	30.0%
Mean	8.82	8.87	9.11	9.07	9.35
Teacher effect ($\tau+r(t)$)					
Mean	-0.011	0.004	0.029	0.036	0.026
SD	[0.151]	[0.155]	[0.134]	[0.154]	[0.139]
Salaries (expressed as multiple of baseline starting salary)					
Base starting salary	1.000	0.964	1.054	0.978	1.059
Average total pay	1.148	1.168	1.216	1.180	1.226
<i>Change from baseline</i>					
Teacher ability (τ)					
Mean		+0.015	+0.041	+0.047	+0.038
SD		+0.003	-0.020	-0.000	-0.015
Teacher experience					
Pct. 1st year		-0.0 p.p.	+0.1 p.p.	-0.1 p.p.	+0.4 p.p.
Pct. 1st 3 years		-0.1 p.p.	+0.1 p.p.	-0.4 p.p.	-0.8 p.p.
Mean		+0.045	+0.284	+0.249	+0.531
Teacher effect ($\tau+r(t)$)					
Mean		+0.015	+0.040	+0.047	+0.037
SD		+0.004	-0.017	+0.002	-0.013
Salaries					
Base starting salary		-3.6%	+5.4%	-2.2%	+5.9%
Average total pay		+1.8%	+5.9%	+2.8%	+6.8%

Table A.2: Pay-for-performance contract with multidimensional output and corruptible performance measure

	Baseline	Multi-dimensional output		
		No influence	With costly influence	With costly & counterproductive influence
	(1)	(2)	(3)	(4)
Key parameters				
corr(τ_1, τ_2)	1	0.4	0.4	0.4
Influence possible?	n	n	y	y
Influence is counterproductive	n	n	n	y
Impact of performance pay policy on:				
Effectiveness on dimension A				
As measured	+0.046	+0.046	+0.067	+0.067
Net of influence	n/a	n/a	+0.039	+0.039
Effectiveness on dimension B	+0.046	+0.026	+0.025	-0.003
Average total pay	+2.9%	+2.9%	+3.5%	+3.5%