Teacher Quality Policy When Supply Matters

Jesse Rothstein*

University of California, Berkeley and NBER

June 2012

Abstract

Many states and school districts, encouraged by the U.S. Department of Education and private philanthropists, are moving to implement new teaching contracts that provide bonuses to effective teachers and/or ease the removal of ineffective teachers. An important intended mechanism is self selection: These policies should make teaching more attractive to those who are effective and less attractive to those who are not. But we know little about the plausible selection effects of such policies. I develop a model of the teacher labor market, incorporating dynamic self selection through entry and voluntary exit, noisy performance measurement, Bayesian learning, risk aversion, and on-the-job search, and I use this model to study the effects of alternative contracts on teaching effectiveness and on the teacher wage bill. Simulations with reasonable parameter values indicate that labor market interactions are important components of the evaluation of each policy. Plausible performance pay policies have only tiny effects on teacher quality. Firing policies can have larger effects, albeit only with substantial increases in teacher salaries. Moreover, benefits are extremely sensitive to the quality of the performance measure – misalignment between measured performance and true productivity substantially attenuates these benefits.

1 Introduction

A 2010 "manifesto" signed by sixteen big-city school superintendents confidently states that "the single most important factor determining whether students succeed in school...is the quality of their teacher" (Klein et al., 2010). This has become the consensus view among researchers, advocates,

*E-mail: rothstein@berkeley.edu. I thank David Card, Sean Corcoran, Richard Rothstein, and conference and seminar participants at APPAM, Berkeley, Northwestern, and NYU for helpful discussions.

and policymakers (Bill & Melinda Gates Foundation, 2010b; Chetty et al., 2011; Hiatt, 2009). Advocates promise enormous benefits from policies aimed at improving teacher quality, sometimes going so far as to say that they can "turn our schools around" (Gates, 2011; see also Hanushek, 2010).

Researchers have made great strides in recent years toward developing and validating measures of teacher quality (McCaffrey et al., 2003; Kane and Staiger, 2008; Bill & Melinda Gates Foundation, 2010a, 2012; Chetty et al., 2011), though important questions remain (Rothstein, 2010, 2011; Corcoran et al., 2011). But relatively little attention has been paid to the design of policies that will use the new measures to raise teacher effectiveness and, thereby, student achievement.

Four broad classes of policies might be considered: (i) better selection for effectiveness on entry into the profession, (ii) professional development aimed at helping teachers learn to be more effective, (iii) differentiation of teacher compensation based on demonstrated effectiveness, and (iv) selective non-retention of teachers who are found to be ineffective. But economists in particular and policymakers in general have tended to focus on the final two types of policies. Secretary of Education Arne Duncan, for example, has said, "We have to reward excellence....We also have to make it easier to get rid of teachers when learning isn't happening" (Hiatt, 2009).

A few recent experiments have examined the short-term effects of performance rewards (Marsh et al., 2011; Goodman and Turner, 2010; Fryer, 2011; Springer et al., 2010), with generally disappointing results. But these experiments aimed at eliciting greater effort from teachers, which may be the wrong margin. Many observers believe that variation in teacher effectiveness primarily reflects largely immutable personality traits.[1] Under this view, neither effort nor professional development is particularly relevant; the primary mechanism by which instructional quality might be improved is through selection.

In principle, a well designed contract could make the profession more attractive to effective teachers and less attractive (or, with firing policies, unavailable) to ineffective teachers (Lazear,

[1]For example, the 2010 superintendents' manifesto urges us to "stop pretending that everyone who goes into the classroom has the ability and temperament" to be an effective teacher (Klein et al., 2010).

2003).[2] The experiments to date are unable to uncover effects operating through this channel. Decisions about entry into teaching and about retention in the profession depend on expected compensation over the entire future career, and a short-term experimental intervention cannot have large effects on this (Hess, 2010). Even quasi-experimental approaches are not very promising. Murnane and Cohen (1986) note that past performance pay systems have invariably been abandoned quickly.[3] Thus, potential teachers are unlikely to expect that more recently introduced policies will persist long enough to generate important variation in lifetime compensation.

This paper takes an alternative approach to understanding the impact of alternative contracts on teacher selection. I develop a stylized model of the performance measurement process and of the teacher labor market, and use simulations of the model to examine the effects of performance pay and selective non-retention on the distribution of teacher quality. The model incorporates private information on the part of teachers about their own ability and Bayesian updating by both the employer and the teacher herself in response to noisy annual performance signals. 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.

My analysis is most closely related to an examination of nonretention policies by Staiger and Rockoff (2010; see also Gordon et al., 2006), who also treat teacher quality as an imperfectly observed trait. In their model, the only cost of firing a poor teacher is that the district will have to hire an inexperienced, but otherwise average, replacement. Not surprisingly, then, Staiger and Rockoff (2010) conclude that that the optimal policy would fire a large share of teachers – as many

[2] An alternative to changing contracts would be to do a better job of selecting for effectiveness in hiring decisions. Researchers have had trouble, however, identifying characteristics that one might observe at the time of hiring that are strongly correlated with subsequent effectiveness (Hanushek and Rivkin, 2006; Clotfelter et al., 2007; Goldhaber and Brewer, 1997), though Rockoff et al. (2011) were more successful. Taylor and Tyler's (2011) examination of a formative evaluation program for experienced teachers found large impacts on teachers' subsequent performance, seemingly falsifying the the immutable trait view.

[3] As Steve Glazerman (as quoted by Sparks, 2010) puts it, "One of the reasons it's been very difficult to evaluate performance pay is the program often doesn't outlast the evaluation."

as 80% – early in their careers and that this would dramatically improve student achievement. But they assume that replacement teachers can be hired without limit, with no increase in compensation, from the same quality distribution from which current teachers are drawn.

This is implausible. The elasticity of labor supply to teaching is unlikely to be infinite. Moreover, one would expect reductions in job security to require a compensating differential even to maintain the current level of supply. Thus, if salaries are not adjusted, filling positions vacated by fired teachers would likely require reducing hiring standards and thus, presumably, quality.[4]

A recognition that alternative teacher contracts must operate within a labor market thus implies that firing policies have important costs that are not incorporated in the Staiger and Rockoff (2010) analysis. It also suggests that performance pay, which could attract high ability people into the profession, may be an attractive alternative or complement. But quantifying the costs and benefits of these policies requires a model of the interaction of teacher contracts with the teacher labor market.

In my model, labor supply to teaching is treated as a dynamic, discrete choice problem under uncertainty. Decisions to enter into teaching depend on risk-adjusted expected compensation over the whole career, with risk deriving both from imprecision in performance measurement and from uncertainty about one's own ability. Similarly, experienced teachers engage in ongoing on-the-job search for non-teaching options, and as a result attrition decisions depend as well on expected future compensation. Alternative contracts affect the future compensation and job security of a teaching job, with differential expected impacts on teachers who vary in their estimates of their own ability. These expectations in turn have differential effects on both decisions to enter the profession and to exit for other opportunities.

In the absence of a compelling setting in which to estimate the model, I instead simulate it

[4]Although 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; Murnane and Steele, 2007); Gordon et al. (2006) wrote around the same time about "the coming teacher shortage." Shortfalls are traditionally filled by hiring teachers with "emergency" credentials, often semi-permanently. Some have hypothesized that current credentialling rules are unrelated to quality and serve primarily as a barrier to entry. If so, loosening of requirements could serve to offset other changes that reduce supply. Of course, changes in entry requirements need not be accompanied by changes in pay or retention policies.

using calibrated parameter values. Key parameters of the performance measurement process and of the teacher labor market are matched to estimates from the literature, though extensive sensitivity checks explore the robustness of the results.

Both differentiated pay and selective nonretention raise average teacher effectiveness but also require increases in average teacher compensation. In the case of nonretention policies, these increases compensate for the cost of possible future job loss and accommodate the need to hire more starting teachers. Under performance pay, which I assume is coupled with a cut to base salaries calibrated to maintain the current size of the teacher workforce, average compensation must rise to compensate for the increase in risk. A natural question is whether these policies are cost effective. To assess this, I compare the alternative contracts to a traditional resource policy, modeled as a change in class sizes. Thus, my analysis reveals whether a marginal dollar is better spent on reducing class sizes or on policies aimed at raising teacher quality, exactly the question that has been raised in recent discussions of education policy.[5]

To focus on the selection effects of performance accountibility policies, I ignore all effects of these policies other than those operating through selection. In my main model, teachers can do nothing to influence their actual or measured performance. As noted above, the evidence to date suggests that alternative contracts do not elicit greater effort, or at least that any such effort does not translate into greater productivity (Springer et al., 2010; Marsh et al., 2011; Fryer, 2011; Goodman and Turner, 2010). On the other hand, there is reason to think that high-stakes performance measurement could lead to distortion of the performance measure (Campbell, 1979), perhaps through narrowing of curricula and redirection of effort toward measured outcomes (Corcoran et al., 2011; Carrell and West, 2010; Rothstein, 2011), through changes in student assignments (Rothstein, 2010, 2009a), or even through outright cheating (Jacob and Levitt, 2003). It could also crowd out intrinsic motivation, thereby *lowering* teacher effort (Kreps, 1997; Jacobson, 1995), or discourage cooperation among coworkers. The potential for manipulation and, worse, goal distortion militates against high-stakes uses of the performance measures (Baker, 1992, 2002;

[5]For example, Gates (2011) proposes coupling an increase in class size with higher salaries for effective teachers.

Holmstrom and Milgrom, 1991).[6] Because I ignore such issues, my primary analysis almost certainly overstates the likely effects of a shift to a performance-sensitive contract. Thus, 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 based on measured output.

In general, I choose parameters to make the best realistic case for alternative contracts. One aspect of my modelling strategy, however, works in the other direction, leading me to understate the potential benefits of certain versions of the policies under consideration. I focus on policies that are implemented at scale, by the entire education sector rather than by an individual district. Labor supply to a district is more elastic than that to the teaching occupation taken as a whole (though even for small districts it is likely to be less than perfectly elastic – see, e.g., Manning, 2005; Ransom and Sims, 2010; Falch, 2010). Thus, even a small performance pay system can enable a district to poach the best performers from its labor market competitors.[7] My interest is in the effects of a systemwide shift, and I thus focus on smaller elasticities relevant to the teaching profession as a whole. Similarly, I rule out the "dance of the lemons," whereby teachers fired for poor performance from one district are rehired by a neighboring district; in my model, a teacher who is fired must leave the profession entirely. This is the right assumption for understanding the kinds of national policy shifts that have been promoted by the Department of Education in recent years, but is likely to understate the impact of a policy change undertaken by a single small district.

2 Relevant Evidence and the Case for Simulations

Rick Hess, responding to the POINT study of teacher performance bonuses (Springer et al., 2010), argues that studies like this cannot measure the selection mechanism:

> Could a randomized field trial be designed that would address the questions we really

[6] It may also imply that some uses are better than others. MacLeod (2003), for example, argues that when worker evaluations are subjective performance pay is is preferable to performance-linked firing policies.

[7] This is one interpretation of the famous Safelite Auto Glass performance pay program (Lazear, 2000). A similar program adopted by the auto glass industry as a whole would likely have had a smaller effect. The ongoing official evaluation of the federal Teacher Incentive Fund will assign individual schools within districts to be eligible or ineligible for incentives (Mathematica Policy Research, undated), severely limiting its ability to identify policy-relevant effects on teacher recruitment.

> care about? In theory, sure. But it would 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)

This understates the difficulty, as one would need somehow to assign the high school students to different possible compensation structures in the event that they decide to become teachers. And even if this could be accomplished, this study "wouldn't tell us what to do today [and] wouldn't generate much in the way of findings until the 2020s" (Hess, 2010).

This challenge motivates my strategy of turning to a structural model as a source of improved understanding. In principle such a 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. Four issues arise in the literature on teacher salaries that are important to the questions at hand.

The first is the distinction between the effects of changes in an individual district's salary schedule and the effects of an across-the-board change in salaries. Most studies (e.g., Figlio, 1997; Guarino et al., 2011; Hanushek et al., 1999; Murnane et al., 1989) examine variation in salaries across nearby districts. Jones and Hartney (2011), for example, find that performance pay implemented at the district level leads to improved teacher recruitment. But as noted above, the margin of interest here is selection into the profession rather than shifts among competing districts. Only a few studies (e.g., Hanushek and Pace, 1995) are informative about the effect of broad-scale changes in salaries. Ransom and Sims (2010) are more careful than most to distinguish the two types of effects. They find that the elasticity of teacher attrition in a moderate-sized school district with respect to that district's relative salary – an upper bound to the elasticity with respect to across-the-board salary changes – is roughly 3 (see also Falch, 2010).

A second distinction is between decisions to enter the profession and decisions to exit for other careers. Some studies focus on teacher recruitment (e.g., Hanushek and Pace, 1995), while others focus on retention and mobility (e.g., Hanushek et al., 1999, 2004). I take the Ransom and Sims (2010) estimate as a reasonable upper bound to the attrition elasticity. In my baseline estimates, I assume that the the entry elasticity is the same – for a total labor supply elasticity of nearly 6 – though I also explore specifications that allow for differential sensitivity on the two margins and for more realistic (i.e., lower) values.

A third important distinction is between the current salary and the expected future salary. Older workers generally earn more than younger workers, both in teaching and in other occupations, and Zarkin (1985) finds that decisions to enter teaching are influenced by expectations of future earnings. As many performance pay and retention policies change the effective returns to experience in teaching and the probability of remaining in the profession for the long term, evaluation of these policies requires an understanding of the life cycle dynamics. In modelling alternative contracts, I assume that teachers' decisions depend on discounted, risk-adjusted future earnings.[8] To abstract from transition issues as teachers hired under one salary schedule are moved to a new one, as well as from attenuation due to uncertainty about future regime changes, I focus on the steady state operation of contracts that are assumed by all agents to be permanent.

Finally, it is useful to distinguish between supply and demand responses. Unlike in competitive markets, teacher salaries do not adjust immediately to equilibrate supply with demand. Frequently, both the number of teachers demanded and the wage are fixed by law or collective bargaining contract. A given wage may produce too many or too few applicants for the number of jobs available, and districts respond in the only way they can, by changing the job requirements. Thus, a district facing a shortfall of applicants will hire uncredentialed teachers who would not be considered by a district with enough credentialed applicants to fill its positions.

[8] One possibility that I do not model is that teaching experience may not be valued in non-teaching jobs. If so – if experienced teachers who move into other occupations are paid like inexperienced workers – then policies that raise the rate at which teachers are displaced may dramatically lower the expected lifetime returns to entering the teaching profession. Unfortunately, I am aware of no good evidence on this.

In a market like this, the main consequence of a change in offered salaries may be to alter the *types* of applicants hired. Some studies find that salary increases lead to increases in the qualifications (e.g., college selectivity or subject matter expertise) of teachers hired (Figlio, 2002), but others find a weak relationship or none at all between salaries and the quality of teaching (Ballou and Podgursky, 1996; Hanushek et al., 1999). Ballou (1996) posits that salaries are weakly linked to quality in the cross section because districts do a poor job of selecting among their applicants. This could well be consistent with an aggregate supply-side effect of salaries on the quality of applicants, as indicated by the coincidence of long-run declines in teacher quality with improvements in non-teaching options for high ability women (Corcoran et al., 2004a,b; Hoxby and Leigh, 2004).

To emphasize the case for differentiated pay, I assume that the supply of teachers of a particular ability type depends on the average risk adjusted compensation that teachers of that type expect, but that districts are unable to distinguish an applicant's ability at the time of hiring. This implies that across-the-board salary increases will induce more applicants but will not raise quality as districts will hire at random from among them. However, changes in the relative compensation of teachers of different ability types can raise quality by shifting the ability distribution in the applicant pool or by inducing differential attrition. Given a particular compensation structure, I assume that base salaries must be set high enough to induce as much total labor supply as is obtained under the current "single salary" schedule.

3 The Model

In this section, I develop a model of the teacher compensation package and occupational choice that can accommodate performance-based contracts and the likely supply responses they will generate. I assume that each individual has some fixed ability as a teacher, but that neither she nor her employer observes it perfectly. Rather, a common signal of ability is observed each period. Teachers themselves use these signals to supplement their private information about their own ability, while the employer can only condition its decisions on the public signals.

An individual uses her private information to forecast her future earnings and job security under the contract on offer, and decides on this basis whether to enter teaching. After she enters, she updates her forecasts each year based on the performance signal, then draws an outside job offer and – assuming she has not been fired – decides whether to take that offer or to remain in teaching. If she accepts an outside offer, she does not reenter teaching.

The teacher contract has two components: The annual pay, and a retention decision. Both may depend on the sequence of signals to date but not on the teacher's ability or private information conditional on that. Importantly, the employer cannot observe even a signal of the ability of a prospective teacher at the time of application, and thus there is no scope for improving the hiring process by making better choices among the candidates. I make this assumption to focus on the effects of performance-based accountability; if in fact it would be possible to do a better job of evaluating and hiring, this could be done now without implementing a performance-based contract and the ability to do this should not be seen as one of the benefits of adopting such a contract.

I assume that any contract is adopted by all districts in the state or nation. There is thus no scope to arbitrage contract terms by choosing among districts, and the relevant labor supply elasticities are those governing supply to the occupation rather than the much larger elasticities faced by individual firms. Similarly, nonretention decisions are globally binding – a teacher fired from one district cannot be rehired by another.

I develop the model in several parts. First, I define the performance measurement process and the Bayesian learning model. Second, I describe the performance-linked contracts that I consider. Third, I discuss 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, developed in greater detail in the appendix.

3.1 Effectiveness, Performance Measurement, and Learning

Teacher i has true ability τ_i. In the current pool of teachers ability is normally distributed with mean zero and standard deviation σ_τ, though changes in the teacher contract may change the selection process and thereby alter that distribution.

A teacher's productivity depends on her ability and her experience, t, with known return-to-experience function $r(t)$. Each year, a noisy signal of the teacher's ability is observed by both the teacher and the employer:[9]

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

The noise component, ε, is normally distributed with mean 0 and standard deviation σ_ε and is independent across both i and t. This implies that there is no bias in the performance measure – there are no teachers who by virtue of the types of students they teach or the methods they use can expect to draw their εs from a different distribution.

Prospective teachers know something of their own personalities, and thus have guesses – prior means – about their abilities as teachers. These guesses, μ, have mean 0 and standard deviation σ_μ in the population of current teachers. I assume that individuals are rational and unbiased in their self assessments.[10] Thus, $\tau \mid \mu \sim \mathbb{N}\left(\mu, \sigma_\tau^2 - \sigma_\mu^2\right)$, and a prospective teacher with signal μ uses this as her prior distribution for her own ability. The precision of potential teachers' guesses can be measured as $h \equiv V(E[\tau\mid\mu])/V(\tau) = \sigma_\mu^2/\sigma_\tau^2$, where $h = 1$ corresponds to perfect accuracy and $h = 0$ to a total lack of information.

Teachers update their priors rationally as they accumulate performance signals. After t years, a teacher's posterior is

$$\tau \mid \{\mu, y_1, \ldots, y_t\} \sim \mathbb{N}\left(\frac{t^{-1}\sigma_\varepsilon^2\mu + (1-h)\,\sigma_\tau^2\bar{y}_t}{t^{-1}\sigma_\varepsilon^2 + (1-h)\,\sigma_\tau^2}, \frac{1}{t\sigma_\varepsilon^{-2} + (1-h)^{-1}\,\sigma_\tau^{-2}}\right), \tag{2}$$

where $\bar{y}_t \equiv t^{-1}\sum_{s=1}^{t} y_s$ is the average performance signal to date. I denote the teacher's posterior mean after t years of experience 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.

[9] In practice, the available performance measures confound ability and experience – the observed signal is $y_{it} + r(t)$. But so long as the $r()$ function is common knowledge, all parties can easily back out y_{it}.

[10] Prospective teachers may overestimate their own effectiveness, perhaps particularly so when that effectiveness is low. This would dilute the effect of performance incentives on recruitment, as even bad teachers would respond to incentives meant for good ones.

Employers are unable to observe μ. Thus, on entry the employer's prior is uninformative: $\tau \sim \mathbb{N}\left(0, \sigma_\tau^2\right)$, while after t years the posterior ability distribution is

$$\tau|\left\{y_1, \ldots, y_t\right\} \sim \mathbb{N}\left(\frac{\sigma_\tau^2}{t^{-1}\sigma_\varepsilon^2 + \sigma_\tau^2}\bar{y}_t, \frac{1}{t\sigma_\varepsilon^{-2} + \sigma_\tau^{-2}}\right). \tag{3}$$

3.2 Alternative contracts

In the baseline contract, all teachers are retained every year, though they may depart voluntarily. Pay rises with t but is insensitive to y_{it}: $w_{it}^0 = w^0\left(1 + g\left(t\right)\right)$, with $g'\left(\right) > 0$. Any teacher still present after T years exits with probability one at that point.

Alternative contracts base either the compensation or the retention decision on the sequence of performance signals to date. Under a performance pay contract, pay is

$$w_{it}^{PP} = \alpha^{PP} w^0\left(1 + g\left(t\right)\right)\left(1 + b * e\left(y_{i1}, \ldots, y_{it}; t\right)\right), \tag{4}$$

where $e\left(\right)$ is an indicator for bonus receipt, b is the size of the bonus, expressed as a fraction of base pay, and $\alpha^{PP} < 1$ is an adjustment to the base salary calibrated to yield an appropriate number of teachers. Receipt of a bonus can depend only on observed performance signals. I examine a bonus that is based on average performance over the previous two years:

$$e\left(y_{i1}, \ldots, y_{it}; t\right) = \begin{cases} 0 & \text{if } t = 1 \\ \mathbb{1}\left(\frac{y_{it} + y_{i,t-1}}{2} > \bar{y}^{PP}\right) & \text{if } t > 1 \end{cases} \tag{5}$$

Under a firing contract, a binary decision is made each year about whether to retain a teacher. Teachers who are retained are paid according to the single salary schedule, $w_{it}^F = \alpha^F w_{it}^0$. As with the performance pay contract, I calibrate the base salary to attract enough teachers to fill the available positions. Under the firing contract, this requires a premium over the existing salary schedule, $\alpha^F > 1$, both because more teachers must be hired and because teachers demand compensation for the increased risk.

Many recent policy discussions have focused on once-and-for-all retention decisions made early in the career: A teacher is evaluated after, say, two years, and is not retained if her performance fails to exceed a specified threshold, but once she passes that point she is tenured and can no longer be fired except in extreme circumstances (e.g., Gordon et al., 2006; Staiger and Rockoff, 2010). This sort of contract leaves a great deal of information on the table – why should a teacher who squeaks past the threshold in the first two years but then is later revealed to be of very low quality not be let go at that point? I thus focus on a decision rule that allows a teacher to be fired whenever the information available to date indicates that she is of low quality. That is, any teacher for whom the district's posterior mean $\frac{\sigma_\tau^2}{t^{-1}\sigma_\varepsilon^2+\sigma_\tau^2}\bar{y}_t$ falls below a specified threshold $\bar{y}^F$ is released the next year.[11]

Of course, a variety of alternative policies are possible – pay could depend continuously on performance, firing policies could take the form of once-and-for-all tenure decisions, and the two types of policies could be combined.[12] I explore several variants below, in Section 4.3.

3.3 The Teacher Labor Market: Entry and Persistence

A prospective teacher's prior about her own ability generates a prior distribution for the sequence of future performance signals $\{y_1, y_2, \dots\}$ that she will receive if she enters (and stays in) the profession. In combination with a specification of the mapping from performance signals to compensation – as specified by the contract in place – this in turn generates a prior distribution over compensation in each future year that she might remain in the profession, $\{w_1, w_2, \dots\}$. Prospective teachers discount future income rationally with discount rate δ. They are also risk averse with

[11] I defer to future work the question of whether this is the optimal decision rule. I conjecture that the threshold $\bar{y}^F$ should rise with experience – that firing teachers early in their careers creates costly randomness, and that under the optimal policy we should be hesitant to fire a teacher while the posterior variance about her ability remains large. Note, however, that there is an element of this built into the contract specified here: The posterior mean shrinks average performance to date toward zero, with more shrinkage the lower is t. Thus, for low t $\bar{y}_t$ has to be quite low in order for the posterior mean to fall below $\bar{y}^F$, while a teacher with more experience can be fired with a more moderate $\bar{y}_t$.

[12] For the reasons outlined above, even a discontinuous performance bonus should depend on the posterior mean rather than on only the two most recent performance measures. But this would mean that bonuses would be concentrated among highly experienced teachers and, for many, would be invariant to recent performance. Indeed, in many models the optimal bonus would be awarded only upon retirement and be based on average performance over the whole career. Such a bonus system seems unlikely to be palatable, albeit for reasons outside my model.

constant relative relative risk aversion coefficient ρ.

For the moment, set aside both firing and voluntary attrition, and assume that all teaching careers last for exactly T periods. Then the utility that a prospective teacher with private information μ anticipates from a teaching career, scaled in terms of the "certainty equivalent," the dollar value of the lump-sum period-0 payment that would provide the same utility as the uncertain career compensation, is[13]

$$V_0(\mu_i) \equiv \left(E\left[\left(\sum_{t=1}^{T} \delta^t w_t \right)^{\frac{1}{1-\rho}} \middle| \mu_i \right] \right)^{1-\rho}. \tag{6}$$

The inner expectation can be decomposed into the expected compensation given τ_i and a prior distribution over τ_i:

$$V_0(\mu_i) = \left(E\left[E\left[\left(\sum_{t=1}^{T} \delta^t w_t \right)^{\frac{1}{1-\rho}} \middle| \tau_i \right] \middle| \mu_i \right] \right)^{1-\rho}. \tag{7}$$

Decisions to enter the teaching profession depend on $V_0(\mu)$ with elasticity η. Under the baseline contract, w_t is independent of τ (and therefore of μ), so (7) resolves to the present discounted value of the salary stream, $V_0 = \sum_{t=1}^{T} \delta^t w_t^0$. As a consequence, a 1% increase in base pay w^0 raises $V_0(\mu)$ by 1% for all μ and induces a η% increase in the number of applicants of each type. Under a performance pay contract, the distribution of w varies with τ. Thus, the value anticipated by a prospective teacher depends on her private information, μ. Supplementing the base contract with a performance pay program totalling 1% of the baseline salary pool will induce a greater-than-η% increase in high-μ applicants and a smaller increase in low-μ applicants (who anticipate that they will nevertheless receive bonuses with positive probability), while a bonus program that is paid for by reducing base salaries will induce more high-μ and fewer low-μ applications. In each case, the precise numbers depend on the precision of applicants' priors (i.e., on

[13]Equation (6) and the following assumes $\rho \neq 1$. If $\rho = 1$, $V_0(\mu_i) = \exp\left(E\left[\ln\left(\sum_{t=1}^{T} \delta^t w_t\right) \middle| \mu_i\right]\right)$.

h) and on the accuracy with which bonuses are restricted to high-τ teachers. In general, however, if prospective teachers are risk averse the average of V_0 will be less than the average expected discounted future wage, so a budget-neutral introduction of performance pay will lead to reductions in the total number of applicants.

Exit decisions, like entry decisions, depend on the discounted, risk-adjusted expectation of future compensation, this time conditional on both the initial ability signal μ and the sequence of performance measures to date. Computing the actual value function V_t of a teacher considering exiting in year t requires accounting for the option value of remaining in teaching to collect additional performance signals in later years. I thus model the exit decision as a dynamic programming problem for a teacher who engages in on-the-job search and draws a single outside offer each year. Outside offers arrive after the previous year's performance measure and any performance-dependent compensation.

A teacher who receives outside offer v after period $t-1$ will accept the outside offer – and exit teaching permanently before the start of period t – if it offers continuation value $v > V_{it}$, where V_{it} is the continuation value of remaining in teaching. Otherwise, she will remain for at least one more year. This means that the continuation value for someone still in teaching in period t evolves according to:

$$V_{it} = \left(E\left[\left(w_{it} + \delta \max\left(v_{i,t+1}, V_{i,t+1}\right)\right)^{\frac{1}{1-\rho}} \middle| \mu_i, y_{i1}, \ldots, y_{i,t-1}\right]\right)^{1-\rho}. \tag{8}$$

V_{it} depends on the performance signal $y_{i,t-1}$ in three ways: First, w_{it} itself may incorporate a performance bonus that depends in part on the previous year's performance. Second, a teacher who performs poorly may be fired, affecting $V_{i,t+1}$. Third, the realization of y_{it} will influence the teacher's posterior $\hat{\tau}_{it}$, which in turn will influence $V_{i,t+1}$ through her evaluation of her likelihood of receiving future bonuses or of being fired in periods $t+1$ and beyond.

Because the baseline contract incorporates neither bonuses nor firing, V_{it} is unaffected by τ_i under this contract and as a result is constant across teachers. I denote the continuation value under this contract as V_t^0. Under a contract that incorporates firing, a teacher who is fired after period t receives continuation value $V_{i,t+1} = (1-\kappa)V_{t+1}^0$.

I assume that outside offers have continuation value that is drawn from a censored Pareto distribution, calibrated to yield an annual exit hazard under the baseline contract of λ_0, an elasticity of this hazard with respect to a certain, permanent increase in base pay (i.e., to w^0) of $-\zeta$, and a finite V_{it} for any value of ζ.[14] The specific calibration is discussed in the Appendix.

Equation (8) does not have a closed-form solution. I discuss a solution algorithm in the Appendix, relying on recursion to evaluate the value function at each point in the state space $I_t \equiv \{\mu_t, y_1, \ldots, y_{t-1}\}$. This algorithm gives the value V_{it} as a function of I_t under each potential contract. To simulate the impact of these contracts, I draw teachers from the joint distribution of $\{\mu, \tau\}$, then draw performance measures $\{y_1, \ldots, y_T\}$ for each. For teacher, I use the solved dynamic model to compute V_t at each year t under each alternative contract, and use these to compute the effects of alternative contracts 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 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 the worker of being fired. I assume that the continuation value obtained by a worker who is fired after year t equals $(1-\kappa)$ times the continuation value obtained by a retained teacher who is confident that she will never be fired. This is meant to capture the empirical fact that workers who lose their jobs see long-run earnings declines (von Wachter et al., 2009; Davis and von Wachter, 2011). The firing penalty κ is not paid by someone who voluntary exits the profession in advance of being fired. Thus, if κ is large, teachers who anticipate a high probability that they will eventually be

[14] As $T \to \infty$, the expected career length approaches $1/\lambda_0$, so the elasticity of the career length with respect to w^0 approaches ζ. For finite T, this elasticity will be less than ζ. With the baseline parameters used below ($T = 30$, $\lambda^0 = 0.08$, and $\zeta = 3$), the elasticity is roughly $0.77\zeta = 2.3$. The use of a censored rather than an uncensored distribution is a convenience that ensures that V has a finite mean even when ζ is small; I choose a censoring point higher than any possible inside compensation to ensure that the censoring has no impact on the results.

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	0.4
	δ	Discount rate (real)	0.97
	ρ	Coefficient of relative risk aversion	3
	η	Elasticity of entry with respect to base salary 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
Base contract			
	$g(x)$	Return to experience (real)	$0.015 * x$
Performance pay contract			
	b	Bonus size (as share of base pay)	20%
	f^{PP}	Fr. of current teachers who would receive bonus	25%
	α^{PP}	Base pay as share of pay under baseline contract	96.8%
Firing contract			
	f^F	Fr. of current teachers who would be fired next yr.	10%
	α^F	Base pay as share of pay under baseline contract	106.8%
	κ	Effect of being fired	10%

fired will be unlikely to enter the profession and, having entered, will exit at high rates.

3.4 Calibration

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

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; Aaronson et al., 2007; 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) review of the evidence on teacher experience.[15] A number of papers also examine the year-to-year correlation of value-added measures (Sass, 2008; McCaffrey et al., 2009; Bill & Melinda Gates Foundation, 2010a). 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).

There are comparably few good estimates of the key teacher preference parameters. The discount rate and risk aversion parameter are relatively standard, and an 8% annual exit hazard is roughly consistent with the observed national data; see Appendix Figure 1. Section 2 discusses the literature regarding the labor supply elasticity for teaching. 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 likely overstates the elasticity of exit from the profession. I arbitrarily assume that the entry elasticity η is the same (in absolute value). [16]

A key parameter, and the one about which there is the least available information, is h, the precision of prospective teachers' private information. 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 factor analyses of a number of academic and personality characteristics to identify cognitive and non-cognitive components of teacher quality. Their analysis yields 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. An alternative source of information is the data from the POINT study (Springer et al., 2010), participants in which were asked to forecast their probability of winning a performance award. These forecasts — from experienced teachers who certainly had more information about their own effectiveness than would an entering teacher — were uncorrelated with actual award receipt. This

[15] Wiswall (2011) argues that the leveling off of experience effects is an artifact of misspecification, and his estimates indicate substantial increases in effectiveness beyond the third year.

[16] Both elasticities are expressed in terms the inside wage (i.e., to w^0). Because teachers anticipate spending part of their careers outside of teaching, the value function rises less than one-for-one with the inside wage. Thus, calibrating the model requires larger elasticities with respect to V that vary with t.

strongly suggests that h is quite small. My assumption that $h = 0.25$ almost certainly overstates the true value.

I assume that the single salary contract provides for a 1.5% (real) increase for each year of experience. Both alternative contracts are based on this contract and provide similar experience premia. My performance pay contract provides a 20% bonus for teachers whose two-year moving average performance exceeds a fixed threshold $\bar{y}^{PP} = 0.178$. This threshold is set to ensure that 25% of the current teaching workforce would get bonuses (though the threshold is fixed – if the alternative contract attracts more high-τ teachers then a larger share would receive bonuses). The firing contract is calibrated so that whenever the district's posterior mean for a teacher falls below a (similarly fixed) threshold $\bar{y}^{F} = -0.159$ that teacher is fired. This threshold is chosen to ensure that 10% of current teachers would be fired next year, though in steady state the 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 ability distribution for entering teachers, 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.

A teacher who is fired loses $\kappa = 10\%$ of her continuation value. 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.[17]

The final parameters are the adjustments to base pay under the pay-for-performance and firing contracts, α^{PP} and α^{F}. These are calibrated, given the other parameters, to ensure the same total number of teachers (in steady state) as are obtained under the baseline contract. Performance pay raises compensation for a fraction of teachers, so would yield more total teachers unless salaries are lowered for teachers who do not receive bonuses. Thus, $\alpha^{PP} < 1$. Similarly, a

[17]The effects of mass layoffs may overstate the effect if layoffs are disproportionately 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, there is little signal about any individual worker's ability in a mass layoff; future employers may react more negatively to the information that a job candidate was fired from a previous position for poor performance.

firing policy both requires that teachers be hired to replace those who are dismissed and requires that the added risk be offset, so $\alpha^F > 1$. Derivation of the specific values in Table 1 is discussed below.

4 Results

4.1 Strength of incentives

It is helpful to begin by exploring the effects of noise in the performance measure and uncertainty in prospective teachers' estimates of their own ability on the size of the effective incentives created by the alternative contracts. I do so by examining the steps connecting true ability to the perceived incentives created by a performance pay contract. 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 receipt of a performance bonus. For bonuses more than one year in the future ($s > t+1$), the rule of iterated expectations can be used to express this as:

$$E\left[w_{is} \mid \hat{\tau}_{it}\right] = E\left[E\left[E\left[w_{is} | y_{is}, y_{i,s-1}\right] \middle| \tau_i\right] \middle| \hat{\tau}_{it}\right]. \tag{9}$$

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

Figures 1 through 3 show the different components of this decomposition. In Figure 1, I plot the probability that a teacher at each ability (τ) level will win a bonus in a given year – that is, $E\left[e_{is} \mid \tau_i\right] = E\left[E\left[e_{is} | y_{is}, y_{i,s-1}\right] \middle| \tau_i\right]$. In this figure and in the ones 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 $\sigma_\tau = 0.15$).[18] I graph bonus probabilities under four

[18]Of course, under alternative policies this distribution would change. The percentile scores are simply a convenient

mechanisms for assigning bonuses: Based on true ability, on a single performance measure, on the average of two annual performance measures, or on the average of ten annual measures. The first mechanism is infeasible, of course, but shows what one would like to achieve with a performance bonus system. The second series shows that bonuses awarded based on a single performance measure are extremely noisy: A teacher at the 95th percentile of true ability (nearly half a standard deviation above the bonus threshold) has only a 65% chance of receiving a bonus, while a teacher at the 75th percentile (over half a standard deviation below the threshold) has a 34% chance and one at the 50th percentile (over a full SD below the threshold) has a 17% chance. Use of a two-year moving average for bonus determination smooths out much but not all of this noise: A 95th percentile teacher's probability of receiving a bonus rises to 70% while the 75th percentile teacher's probability falls to 28% and the 50th percentile teacher's to 9%. The final series shows that if performance measures are averaged over ten years, noise is essentially eliminated for teachers far from the threshold but there are still frequent misclassifications of teachers closer to it.

Figure 1 indicates that if prospective teachers knew their own ability a performance bonus program based on two years of data would create reasonably strong incentives for high ability individuals to enter teaching, with limited but nontrivial spillover to teachers of very low ability. However, uncertainty among prospective teachers substantially attenuates those incentives. Figure 2 shows the probability that a teacher with various prior means has actual ability above the bonus threshold – that is, $E\left[\mathbb{1}\left(\tau > \bar{y}^{PP}\right) \mid \hat{\tau}_{it}\right]$ – at entry and at various points in the career. Uncertainty about one's own ability is extremely important at the beginning of the career: Only 3% of new teachers think that they have a better-than-even chance of having true ability above the bonus threshold.

As teachers accumulate information over the course of their careers, they quickly learn their places in the distribution. The standard deviation of teachers' posterior means rises from 0.09 at entry to 0.11 after one year of experience, 0.12 after two years, and 0.14 after 10 years, while uncertainty about one's own ability falls commensurately. Thus, after one year of experience, 6%

scale that avoids attracting undue visual attention to the extreme tails of the distribution.

of teachers think they have at least a 50% chance of having a true ability above the bonus threshold, while after two years this rises to 7% and after ten years to 10%.

Figure 3 shows $E\left[e_{is}\,|\,\hat{\tau}_{it}\right]$, reflecting the combined roles of uncertainty about ability and noise in the performance measures. It shows that even a teacher at the 90th percentile of the private information distribution on entry thinks she has only a 37% chance of receiving a bonus in any given year of her career. After one year of experience, a teacher with posterior mean at the 90th percentile thinks her chance of receiving a bonus is 42%, and this rises to 45% after two years and 51% after 10 years.

Equation (9) characterizes the mapping between *perceived* ability and future pay, as graphed in Figure 3. That figure suggests that the bonus program will do a modestly good job of incentivizing teachers who think they are of very high ability, particularly after a few years of experience. But teachers' perceptions are not perfectly accurate: Some low-ability teachers have high priors, and vice versa. It will do little good for these teachers to react to the incentives created by the bonus system. The key question for the efficacy of the performance pay system is whether teachers who actually are of high ability perceive their future pay to have risen. That is, the relevant function for evaluating the strength of the incentives that the performance pay system creates to attract good teachers is $E\left[E\left[w_{is}\,|\,\hat{\tau}_{it}\right]|\,\tau\right]$. The slope of this function is necessarily flatter than is $E\left[w_{is}\,|\,\hat{\tau}_{it}\right]$, simply because $E\left[\,\hat{\tau}_{it}\,|\,\tau\right]$ has a slope less than one.

Figure 4 shows the average anticipated probability of winning a bonus, $E\left[E\left[e_{is}\,|\,\hat{\tau}_{it}\right]|\,\tau_i\right]$, by percentile of *true* ability. It shows that at entry there is very little differentiation except at the extreme tails of the distribution: Even prospective teachers at the 90th percentile of the true ability distribution think that they have only a 30% chance of winning a bonus in any year of their careers, on average, about triple the probability anticipated by the average 10th percentile prospective teacher. Perceived incentives become better targeted as teachers gain experience. After ten years, 90th percentile teachers think they will win bonuses in nearly half of the remaining years of their careers, on average, while 10th percentile teachers anticipate less than a 1% chance of receiving a bonus each year. Thus, while incentive effects of a performance pay system are likely to be weak at the recruitment stage, later attrition decisions may be more sensitive to these incentives.

There is a close, albeit imperfect, mapping from the perceived probability of winning a bonus to the value function (7). Figure 5 shows average continuation values for teachers at different percentiles of the prior ability distribution at four points in the career, while Figure 6 shows these continuation values as functions of *true* ability.[19] In each case, values are expressed as percentages of the value that a teacher at the same point in the career would obtain under the baseline, non-performance-dependent contract. Note that the two figures' vertical axes use different scales – as before, mapping subjective ability estimates back to true ability flattens the curves substantially. Thus, while the performance bonus program raises the anticipated continuation value of a teacher at the 95th percentile of the private information (μ) distribution by 1.7% at career entry and of the teacher at the 95th percentile of the posterior distribution after ten years ($\hat{\tau}_{i,10}$) by 5.0%, the corresponding figures for the 95th percentile of the true ability (τ) distribution are only 0.7% at entry and 1.5% after ten years. These small changes suggest that any behavioral responses to the bonus program will be quite modest, even with relatively large labor supply elasticities.

As Figure 1 illustrates, a big source of slippage in the performance pay 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 evaluation, may be more effective. I thus turn to it next.

The solid line in Figure 7 shows the probability that a teacher at each ability percentile will be fired at some point over a 30 year career, assuming that no one quits voluntarily. This can be compared to the dotted line in Figure 1, which shows the probability that a teacher will be recognized with a performance bonus. Clearly, the firing policy does a much better job of discriminating between good and bad teachers. It is quite successful at identifying the worst teachers: 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. Simultaneously it *avoids* identifying

[19] Average values for the lowest ability teachers are lower than in the baseline contract because I assume that base pay will be reduced with the introduction of the bonus.

average teachers as exceptional – a median teacher has a 9% chance of receiving a bonus but only a 4% chance of ever being fired.

The firing policy is like the performance pay policy, however, in that the incentives created by each are attenuated by teachers' uncertainty about their own abilities. The dashed line in Figure 7 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. This is analogous to the solid line in Figure 4; like it, 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 effective.

Figure 8 shows continuation values under the firing contract. This is more steeply sloped than under the performance pay contract (Figure 6), but even the firing contract does only a modest job of differentiating between high- and low-ability teachers: The range of continuation values is about five percentage points at entry and grows to about seven percentage points later. (Figure 8 shows very low continuation values for experienced teachers at the very bottom of the ability distribution, but these are mostly irrelevant – very few such teachers will attain high levels of experience before being fired.)

4.2 Impact of incentives

Figure 6 indicates that the performance pay contract creates modest incentives to encourage highly effective teachers to enter and remain in teaching, while Figure 8 shows somewhat larger but still modest incentives under the firing contract – 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 9 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. But the firing contract has much larger effects here than does the performance pay contract, reflecting the closer connection of decisions to true ability under the former. Under the firing contract, recruitment of teachers from the top ability quartile grows by 10-20%, reflecting the substantial increases in base pay needed to fill the classrooms vacated by fired teachers, while the number of very low ability recruits falls by a somewhat smaller amount.

Recall from the discussion above that the performance pay contract creates stronger effective incentives for experienced teachers – who have better information about their own abilities – than it does for new entrants. This suggests that the contract may have larger effects on retention than on recruitment. Figure 10 shows average career length under the baseline contract and the two alternatives. The performance pay contract turns out to have only small effects on this margin as well. With 8% annual attrition, the median career lasts only about 8 years. Thus, even relatively strong incentives in later years have only small effects on the overall career length distribution.

The firing contract has much more dramatic effects on career lengths. Average careers of the highest ability teachers grow by nearly two years relative to the baseline contract, due to reduced voluntary attrition resulting from the increase in teacher salaries. Among low ability teachers, career lengths shorten dramatically, to as little as two years on average at the very bottom of the distribution. I also plot in Figure 10 the career lengths that would be seen under the firing contract in the absence of any change in voluntary exit rates. For low ability teachers, this is quite similar to the series that incorporates exit responses, though of course for high ability teachers the beneficial effects of the contract evaporate. Evidently, firing decisions mostly take place too quickly to permit many teachers to self-select out of the profession in advance of an anticipated firing.

Figure 11 presents the combined effects on the entry and exit margins, showing the impact of the two contracts on the steady state number of teachers at each ability level. Not surprisingly, the performance pay contract has relatively small effects, increasing the number of high ability teachers by about 10% or less and reducing the number of low ability teachers by similar amounts. The firing policy is more effective, attracting about 30% more of the highest ability teachers than does the baseline contract while reducing the number of very low ability teachers by as much as 80%. As in Figure 10, much of this low-end effect is the consequence of firing rather than self-selection decisions.

Table 2 shows the effects of the two contracts on teacher ability, teacher experience, teacher effectiveness (combining ability and experience effects), and teacher salaries. As seen in Figure 11, the firing policy has much larger effects on the distribution of teacher ability than does the performance pay contract: The former raises mean ability by 0.047, roughly one-third of a standard deviation, while reducing the variability somewhat, while the latter raises mean ability by only 0.007.

Not surprisingly, the performance pay contract tends to reduce attrition and thereby to raise average experience levels, but the effects are quite small: Under this contract in steady state, the average teacher has about 0.02 years more experience than under the baseline contract, and there are slightly fewer inexperienced teachers. (Note that this understates the effect on attrition of high ability teachers, as this is offset by increased attrition among low ability teachers who gradually learn that they will not qualify for future bonuses.) By contrast, the firing contract leads to increases in the number of inexperienced teachers as bad teachers are quickly replaced. Somewhat more surprisingly, however, it also leads to increases in the average experience of active teachers: Those who aren't fired tend to stay longer, due to the higher salaries.

Total teacher effectiveness reflects both ability and experience effects. These are complementary under a performance pay policy but move in opposite directions under the firing policy. However, Table 2 shows that the effects of the firing policy on the number of inexperienced teachers are too small to have a meaningful effect on overall effectiveness. This contrasts with the result of Staiger and Rockoff (2010), who found that firing policies created real experience tradeoffs. They did not model the supply side of the labor market, and in particular did not consider the base salary increases that would be needed to attract enough teachers to fill the vacancies created by teacher fires. In my model, these increases dramatically reduce the number of vacancies to be filled, by inducing retained teachers to persist longer in the profession, and thus importantly ameliorate the ability-inexperience tradeoff. Thus, Table 2 shows that achievement goes up by the same amount as does teacher ability under each of the two contracts: 0.008 standard deviations under performance pay and 0.047 SDs under the firing policy.

The final rows of Table 2 show the effects of the two policies on teacher salaries. As discussed in Section 3, I adjust base salaries under each policy to yield the same total number of teachers as are employed under the baseline contract, using a numerical search algorithm to compute the necessary adjustment. I find that the performance pay contract will yield too many teachers unless base salaries are reduced by 3.2% (thus, $\alpha^{PP} = 96.8\%$ in Table 1), while the firing contract requires raising base salaries by 6.8% in order to fill all positions. But these are not the only effects on the employer's costs. Under the performance pay contract the bonuses must also be paid, while under each contract the rise in career lengths implies somewhat higher experience premia. Combining these, the model indicates that the performance pay contract would raise the teacher salary bill by 1.8%, while the firing policy would raise it by 7.3%.

These are not trivial changes. One way to scale them is to compute the impact of devoting a similar amount of money to class size reduction. Following Krueger (1999), I assume that class size reductions would require proportional increase in non-salary expenditures and in the number of teachers, and that the attracting the needed additional teachers would require an increase in the base salary. For small x, reducing class sizes by $x\%$ would require hiring $x\%$ more teachers, and thus would require raising wages by $x/\eta_{tot}\%$, where η_{tot} is the total labor supply elasticity, equal to about 5.3 in my model. It would also require an increase of $x\%$ in all non-teacher-salary costs. As teacher salaries represent about 35% of total educational expenditures, the total cost, expressed as a share of the current teacher salary bill, would be roughly $x\left(\frac{1}{share} + \frac{1}{\eta_{tot}}\right)\% \approx 3.03 * x\%$. Using this calculation, class sizes could be reduced by 0.6% for the cost of implementing the performance pay contract and by 2.4% for the cost of the firing contract.

The Tennessee STAR experiment indicated that reducing class size from roughly 22 to roughly 15 raised student achievement by 0.15 standard deviations (Krueger, 1999). Extrapolating under the assumption that the effect is linear in the log of class size, an $x\%$ reduction would raise achievement by about $0.004x$ standard deviations. Thus, class size reduction at the cost of the performance pay program would raise student achievement by 0.002 standard deviations, while the cost of the firing policy could raise achievement by 0.010 standard deviations if devoted instead to class size reduction. Each of the teacher quality policies is thus several times more cost effective than class size reduction, 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 two policies could be expanded while remaining cost effective. Figure 12 shows how the achievement impact and the total cost of the firing policy change as the share of teachers fired (f^F) varies. Firing more than ten percent of teachers leads to larger student achievement effects, but also to rising costs. Costs rise roughly linearly with the firing rate, while the student achievement benefits are concave. Thus, marginal increases in firing rates are most cost effective when starting from a baseline of zero firing, as even a small f^F will lead to the firing of the very worst teachers, and cost effectiveness declines steadily as the firing rate increases. Once f^F rises above about 70%, student achievement starts to decline: At this point, even good teachers are being fired very early, simply because they do not post exceptional performance in their first years, and the cost in reduced teacher experience offsets gains through increasing ability.[20] Importantly, costs become quite large even with much lower firing rates. Setting a retention threshold that would lead to one-third of current teachers being fired would require a 20% increase in the teacher salary bill.

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 under a variety of alternative parameter values. The first row repeats the estimates for the baseline parameter values from Table 2. Rows 2 and 3 vary the amount of private information that prospective teachers have about their own abilities. In row 2, a prospective teacher has the equivalent of one

[20] This replicates Staiger and Rockoff's result that if salary costs are ignored the optimal firing rate is quite high.

annual performance measure, while in row 3 she has the equivalent of two annual signals. Higher levels of private information lead to larger achievement effects, particularly under the performance pay contract. They reduce costs under the firing policy, by reducing the perceived risk of firing among those who in fact are unlikely to be fired, but raise costs slightly under the performance pay contract – under which more bonuses will need to be paid out. Thus, in row 2 the two contracts are approximately equally cost-effective, while in row 3 the performance pay contract dominates.

Row 4 of the Table shows estimates when the annual performance measure is less noisy, with reliability 0.6 in place of the 0.4 used for the earlier results. This has only small effects. Many critics of performance evaluation for teachers have focused on noise in value added measures as a first-order problem. These results suggest that that focus is mistaken, at least within reasonable ranges for the amount of noise.

Rows 5 through 11 show the effects of varying the labor supply elasticities. In general, both policies are more effective when labor supply is more elastic. The cost of a performance pay contract rises with the supply elasticity, while the cost of a firing policy is in general declining in the supply elasticity. Performance pay is not particularly sensitive to the margin on which labor supply adjusts, while firing policies are notably more expensive when exit hazards are elastic than when the same total elasticity is more concentrated on the entry margin.

Finally, rows 12 and 13 show alternative variants on the two contracts, under the baseline parameters. Row 12 shows that when the same bonus payments are allocated as larger bonuses to fewer teachers, the effect on student achievement is halved. Row 13 indicates that a more stringent firing policy leads to larger achievement effects but also to higher costs, as shown previously in Figure 12.

As a final sensitivity analysis, I examine the impact of changes in parameter values on the sensitivity of the results concerning varying firing rates, in Figure 12. Figure 13 repeats the estimates from Figure 12 under three sets of parameters. The solid line shows the baseline parameter values, as in Figure 12. The dashed line shows estimates for relatively inelastic supply ($\eta = \zeta = 1.5$, as in row 6 of Table 3), while the dotted line shows estimates when firing is more costly for the worker ($\kappa = 0.2$). In each case, these changes move the parameters toward more realistic values.

Lower labor supply elasticities reduce the impact of the firing contract on student achievement, though this change is not very large. They also dramatically raise the cost of the policy, particularly at high firing rates. With the lower elasticities, firing policies with high firing rates cease to be more cost effective than is class size reduction.

By contrast, if the cost of displacement is higher than in my baseline analysis, the firing policy becomes *more* effective: The more painful is being fired, the larger is the behavioral response as teachers who think they may be ineffective either avoid entering the profession or exit early to avoid bearing the firing cost. Not surprisingly, the higher firing costs also make the policy more expensive, as teachers demand compensation for the increased risk. Cost-effectiveness falls somewhat, though not nearly as much as under the reduced supply elasticities.

5 Misalignment of Performance Measure & Goal Distortion

I have assumed thus far that the performance measure is a noisy but otherwise perfect measure of teacher productivity. But 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. Thus, a performance measurement system will identify teachers who excel on the dimensions being measured, whether or not they are equally good on unmeasured dimensions. Moreover, there may be scope for teachers to raise their measured performance without improving their overall productivity by redirecting effort from unmeasured to measured dimensions (Baker, 1992, 2002; Holmstrom and Milgrom, 1991).[21] Either will attenuate the effects of performance-based contracts. In this Section, I attempt to explore this attenuation. Results are necessarily extremely speculative, as very little is known

[21] Other forms of influence activities, such as cheating or teaching to the test, will improve measured performance but not reflect output even on the dimensions covered by the performance measure.

about either factor.

Unfortunately, efforts to correlate measured performance with true productivity are severely hampered by the lack of an agreed-upon definition of true, comprehensive productivity. But there is suggestive evidence that this correlation may not be very high. The Gates Foundation's Measures of Effective Teaching (MET) project, for example, has found that teachers' value added 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, 2010a; Rothstein, 2011).[22] The correlation between traditional value added and an even more comprehensive productivity measure is unlikely to be higher than this.

I augment the model developed above to incorporate a second dimension of output. I assume that each teacher performs two tasks and that her ability to perform the first, τ_{iA}, may be 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 of those dimensions, $y_{it} = \tau_{iA} + \varepsilon_{it}$.

The first column of Table 4 presents estimates for the firing policy when the two dimensions of ability are perfectly correlated. These are identical to those seen earlier. In the second column, I assume that $\text{corr}(\tau_{iA}, \tau_{iB}) = 0.4$, consistent with the MET evidence. Because the firing policy selects (imperfectly) only on τ_{iA} and because $E[\tau_{iB} \mid \tau_{iA}] = \text{corr}(\tau_{iB}, \tau_{iA}) * \tau_{iA}$, the effect of the policy on the second dimension of teacher output is only 40% 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 dimension 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

[22] These are "disattenuated" correlations that abstract from year-to-year variation in the same teacher's score on each dimension. That is, they measure the correlation between the teacher's two τs, not between her two ys.

influence activities in teaching.[23] 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 bears 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 salary.[24] This is quite 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 prior about her own ability. In my framework, most teachers find it optimal to distort their measured performance at least a bit early in their careers, but typical distortions are fairly small. Among first year teachers, for example, 85% exert positive effort to influence their measured performance, but the average distortion is $E = 0.034$, less than 14% of a standard deviation of measured performance. Influence activities fall off quickly as it becomes clear which teachers are at serious risk of being fired – by the 3rd year, just over one-third of teachers exert positive effort, and by the 7th year the share is below 20%.

The presence of 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 higher than in the baseline model without manipulation, at +0.050 versus +0.047 in the baseline model. But when the effects of distortionary effort are excluded – which they should be if the influence activity takes an unpro-

[23]Carrell 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 Campbell (1979), Rothstein (2009b), and the studies reviewed in Section 4.2 of Figlio and Loeb (2011).

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

ductive form such as cheating or teaching to the test – the policy's impact is only 87% as large as in the base model. And the effects on true (dimension 2) effectiveness are even smaller, only 0.016 student-level standard deviations.

Even this may understate the degree to which distortionary effort can undercut the policy. In column 3, I assume that manipulation of the performance measure is costly to the teacher but has no negative consequences for students. Many forms of manipulation – e.g., narrowing of the curriculum or diversion of class time to test preparation – will undermine student learning in non-tested dimensions. 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 nearly eliminated: The negative consequences of teaching to the test offset more than half of the already small positive effect of improved selection on τ_{iB} from column 3.

It must be emphasized that the effort model here is entirely ad hoc, 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 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 policies to raise teacher quality 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 student scores.

Assuming that the necessary funding is available and that teachers are unable to game the performance measurement process, both policies appear to be cost effective at modest scales relative to "traditional" uses of additional funds. Indeed, recognition of the labor market effects can makes firing policies even more effective than when these effects are ignored, as the accompanying salary increases help to attract and retain high ability teachers.

There are several important caveats to the results, however. First, and most importantly, my main results 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 are corrupted when the stakes are raised.[25]

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. The firing policy is by far the most effective, but it would raise student achievement by only 0.05 standard deviations – nothing to sneer at, but also not a fundamental change. The performance bonus policy would have a smaller, trivial effect on teacher selection.

These impacts depend importantly on the ability to change the base salary to accommodate them. Without increases in base salaries, districts will have great trouble filling the classrooms va-

[25]Chetty et al. (2011) study the relationship between teachers' value-added and students' later earnings. But their analysis can only show that the correlation between value-added and the outputs that influence later earnings is positive; they do not estimate the magnitude. The MET Project (Bill & Melinda Gates Foundation, 2012) is examining the correlations among different performance measures, but under the maintained assumption that value-added equates to true productivity. The project has not shed light on the extent to which the alternative measures are capturing important dimensions of teacher output that are missed by value-added models.

cated by fired teachers. Performance bonuses could be – and sometimes have been – implemented as "add ons" to existing salary schedules, but this would dramatically increase their cost and they would cease to be cost effective even under my extremely favorable assumptions. This conclusion might be reversed if it were possible to effectively screen teacher applicants for quality at the entry point, but there is little evidence to suggest that this is an option.

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.

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 beginning 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 succesful implementation of a performance measurement system that is not susceptible to manipulation and that is accompanied by substantial increases in the resources devoted to teacher pay.

References

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and student achievement in the Chicago public high schools," *Journal of Labor Economics*, 2007, *24* (1), 95–135.

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.

_ **and Michael Podgursky**, *Teacher pay and teacher quality*, Kalamazoo, MI: WE Upjohn Institute for Employment Research, 1996.

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.

_ , "Working With Teachers to Develop Fair and Reliable Measures of Effective Teaching," White Paper, 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.

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?,” January 2007.

Corcoran, Sean P., Jennifer L. Jennings, and Andrew A. Beveridge, “Teacher effectiveness on high- and low-stakes tests,” April 2011. Unpublished manuscript.

_, 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.

_, _, and _, “Women, the labor market, and the declining relative quality of teachers,” *Journal of Policy Analysis and Management*, 2004, *23* (3), 449–470.

Davis, Steven J. and Till von Wachter, “Recessions and the Costs of Job Loss,” *Brookings Papers on Economic Activity*, Fall 2011, pp. 1–69.

Falch, Torberg, “The elasticity of labor supply at the establishment level,” *Journal of Labor Economics*, April 2010, *28* (2), 237–266.

Figlio, David, “Teacher salaries and teacher quality,” *Economics Letters*, 1997, *55* (2), 267–271.

_, “Can public schools buy better-qualified teachers?,” *Industrial and Labor Relations Review*, 2002, *55* (4), pp. 686–699.

_ 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.

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

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

Goldhaber, Dan and Dominic J. Brewer, “Why Don’t Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity,” *Journal of Human Resources*, 1997, *32* (3), 505–523.

Goodman, Sarena F. and Lesley J. Turner, “Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program,” Manuscript 2010.

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.

Guarino, Cassandra M., Abigail B. Brown, and Adam E. Wyse, "Can districts keep good teachers in the schools that need them most?," *Economics of Education Review*, October 2011, *30* (5), 962–979.

Hanushek, Eric A., "The economic value of higher teacher quality," working paper 16606, National Bureau of Economic Research, December 2010.

_ **and Richard R. Pace**, "Who chooses to teach (and why)?," *Economics of Education Review*, 1995, *14* (2), 101–117.

_ **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.

_ **, John F. Kain, and Steven G. Rivkin**, "Do higher salaries buy better teachers?," working paper 7082, National Bureau of Economic Research, April 1999.

_ **,** _ **, and** _ , "Why public schools lose teachers," *Journal of Human Resources*, 2004, *XXXIX* (2), 326–354.

Hess, Rick, "Missing the POINT: Tomorrow's big merit pay study will tell us...nothing," *Education Week: Rick Hess Straight Up [blog]*, 2010. `http://blogs.edweek.org/edweek/rick_hess_straight_up/2010/09/missing_the_point_tomorrows_big_merit_pay_study_will_tell_us_nothing.html`.

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.

Hoxby, Caroline M. and Andrew Leigh, "Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States," *The American Economic Review*, 2004, *94* (2), 236–240.

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.

Jones, Michael and Michael Hartney, "Show Who the Money? Does Performance Pay Attract Higher Quality Teachers?," November 17 2011.

Kane, Thomas J and Douglas O Staiger, "Are Teacher-Level Value-Added Estimates Biased? An Experimental Validaton of Non-Experimental Estimates," March 2008.

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.

Kreps, David M., "Intrinsic motivation and extrinsic incentives," *The American Economic Review*, 1997, *87* (2), 359–364.

Krueger, Alan B., "Experimental estimates of education production functions," *Quarterly Journal of Economics*, 1999, *114* (2), 497–532.

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.

MacLeod, W. Bentley, "Optimal contracting with subjective evaluation," *The American Economic Review*, March 2003, *93* (1), 216–240.

Manning, Alan, *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton, NJ: Princeton University Press, 2005.

Marsh, Julie A., Matthew G. Springer, Daniel F. McCaffrey, Kun Yuan, Scott Epstein, Julia Koppich, Nidhi Kalra, Catherine DiMartino, and Art Peng, "A Big Apple for Educators: New York City's Experiment with Schoolwide Performance Bonuses: Final Evaluation Report," Technical Report, RAND Corporation, Santa Monica 2011.

Mathematica Policy Research, "Information about the national evaluation of the Teacher Incentive Fund (TIF)," undated. Accessed from `http://mathematica-mpr.org/publications/pdfs/TIFEvaluationInfo.pdfonAugust23,2011`.

McCaffrey, Daniel F., J. R. Lockwood, Daniel M. Koretz, and Laura S. Hamilton, "Evaluating Value-Added Models for Teacher Accountability," Report, RAND 2003.

_ , Tim R. Sass, J.R. Lockwood, and Kata Mihaly, "The intertemporal variability of teacher effect estimates," *Education Finance and Policy*, 2009, *4* (4), 572–606.

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 Jennifer L. Steele, "What Is the Problem? The Challenge of Providing Effective Teachers for All Children," *The Future of Children*, Spring 2007, *17* (1).

_ , Judith D. Singer, and John B. Willett, "The influences of salaries and "opportunity costs" on teachers' career choices: Evidence from North Carolina," *Harvard Educational Review*, 1989, *59* (3), 325–346.

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. `http://nepc.colorado.edu/files/TTR-MET-Rothstein.pdf`.

Rothstein, Richard, "The Influence of Scholarship and Experience in Other Fields on Teacher Compensation Reform," in Matthew G. Springer, ed., *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, 2009, chapter 5, pp. 87–110.

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.

Sparks, Sarah D., "ED sweetens the deal for evaluating Teacher Incentive Fund," *Education Week: Inside School Research [blog]*, September 27 2010. `http://blogs.edweek.org/edweek/inside-school-research/2010/09/ed_proposes_new_teacher_incent.html`.

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 Performance: Evidence from Longitudinal Student Achievement Data of Mid-career Teachers," Working Paper 16877, National Bureau of Economic Research March 2011.

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, "The dynamics of teacher quality," January 1 2011.

Zarkin, Gary A., "Occupational choice: An application to the market for public school teachers," *The Quarterly Journal of Economics*, 1985, *100* (2), 409–446.

A Appendix: On-the-job search and exit from teaching

A.1 Search model

I assume that all teachers are engaged in continuous on-the-job search and that 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, v. I assume that the outside offer v_t has a censored Pareto distribution:

$$F_t(v_t) = \begin{cases} 0 & \text{if } v \leq V_t^0 \lambda_0^{-1/\zeta'} \\ 1 - \lambda_0 \left(\frac{v}{V_t^0}\right)^{\zeta'} & \text{if } V_t^0 \lambda_0^{-1/\zeta'} < v < HV_t^0 \\ 1 & \text{if } HV_t^0 \leq v. \end{cases} \tag{10}$$

Here, λ_0 is the baseline annual exit hazard and H is the maximum outside wage, expressed as a fraction of the inside continuation value. Importantly, the distribution of v_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.

The outside distribution (10) is chosen to yield a constant elasticity of the exit hazard with respect to the inside value: If we define $\lambda_t(V_t) = \Pr\{v_t > V_t\} = \lambda_0 \left(V_t/V_t^0\right)^{\zeta'}$ as the probability that a teacher who would obtain continuation value $V_t \in \left[V_t^0 \lambda_0^{-1/\zeta'}, HV_t^0\right]$ in teaching will instead exit, it is straightforward to show that $\partial \ln \lambda_t(V_t) / \partial \ln V_t = \zeta'$. The use of a censored distribution is a convenience that ensures that the outside wage has a finite mean even when $|\zeta'|$ is small; it has no impact on the results so long as H is set larger than the maximum value that can be obtained as a teacher under the contracts considered. In the estimates presented here, I set $H = 2$; this far exceeds the value of even a certain 25% bonus in each future year.

The model in the main text was developed in terms of the elasticity of the exit hazard with respect to the inside wage under the baseline contract, $\zeta \equiv \partial \ln \lambda_t / \partial \ln w^0 = \partial \ln \lambda_t / \partial \ln V_{it} * \partial \ln V_{it} / \partial \ln w^0 = \zeta_t' * \partial \ln V_{it} / \partial \ln w^0$. With on-the-job search, the latter fraction is less than one and varies with t. I thus solve recursively for this elasticity – which depends on ζ_s', $s > t$, but not on ζ_t' itself – and define the elasticity parameter in (10) as $\zeta_t' \equiv \zeta * \left(\partial \ln V_{it} / \partial \ln w^0\right)^{-1}$. Similarly, the elasticity of teacher entry with respect to the initial value V_{i0} is computed as $\eta' \equiv \eta \left(\partial \ln V_{i0} / \partial \ln w^0\right)^{-1}$.

A.2 Solving the model

Equation (8) 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 the next period's performance measure given $I_t \equiv \{\mu_t, y_1, \ldots, y_{t-1}\}$ is

$$y_t \,|\, I_t \sim \mathbb{N}\left(\hat{\tau}_{t-1}, \frac{1}{\frac{1}{(1-h)\sigma_\tau^2} + \frac{t-1}{\sigma_\varepsilon^2}} + \sigma_\varepsilon^2\right). \tag{11}$$

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 entirely trivial.

The recursive solution thus has three steps. First, I compute $w_T^C(y_1, \ldots, y_T)$, 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^C, as a function of I_T, by integrating w_T^C over the conditional distribution of y_T given by (11). Third, for each $t < T$, given estimates of V_{t+1}^C as a function of I_{t+1}, I compute $w_t^C(y_1, \ldots, y_t)$ for each possible $(y_1, \ldots, y_t)$, then integrate over the distribution of y_t (and therefore of I_{t+1}) given I_t to obtain V_t^C.

The state space I_t is of dimension t, creating a curse of dimensionality 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 $\{y_1, \ldots, y_t\}$ 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 $\hat{\tau}_{i,t-1}$ or by the pair $\{\mu, \bar{y}_{t-1}\}$. I can thus focus on state spaces of only three dimensions, $\{\hat{\tau}_{t-1}, y_{t-1}, y_t\}$ for the performance pay contract or $\{\mu, \bar{y}_{t-1}, y_t\}$ for the firing contract. I approximate each of these with grids of 149^3 points.[26]

Having solved for the value functions at each period, I simulate the impact of policies by drawing potential teachers from the joint distribution of $\{\mu, \tau\}$, then drawing performance measures $\{y_1, \ldots, y_T\}$ for each. For each career, I compute I_t and V_t at each year t, and use these to compute the effects of alternative contracts 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 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-$\hat{\tau}_t$ teachers, without offsetting reductions from teachers with low μ or $\hat{\tau}_t$. Under each alternative contract, I compute the steady-state size of the teacher

[26] In 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.

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^{PP} 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.

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

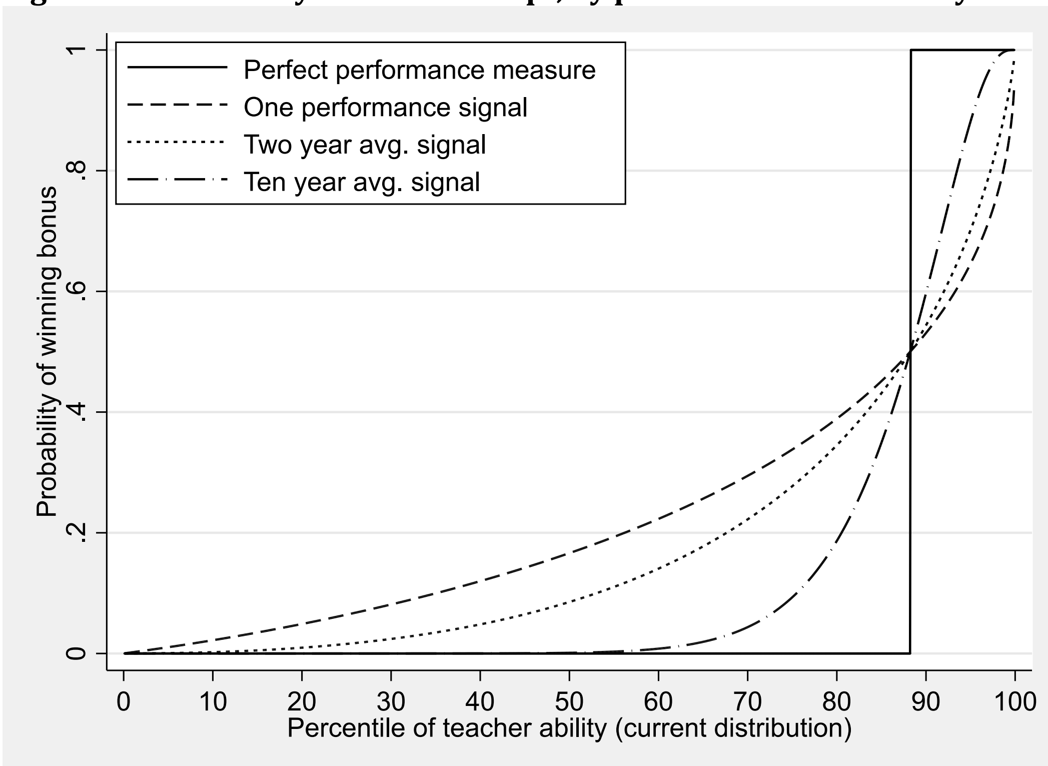

Figure 2. Probability that true ability exceeds bonus threshold, by prior percentile and years of experience

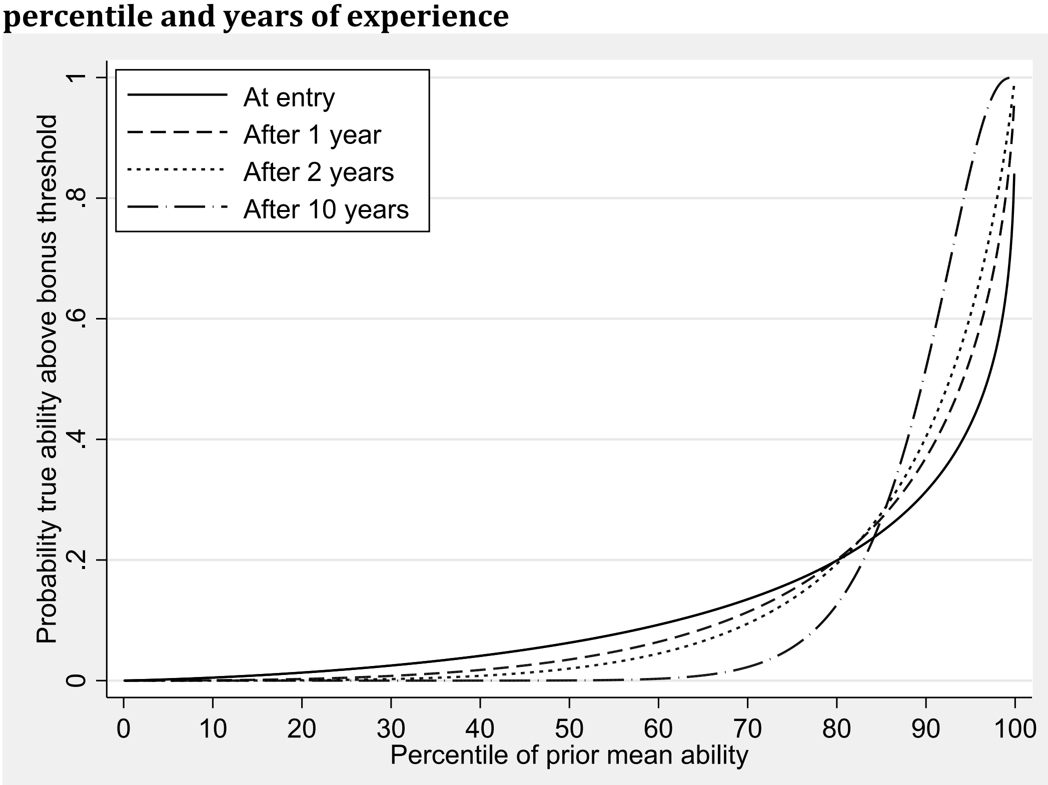

Figure 3. Probability of bonus receipt, by prior percentile and years of experience

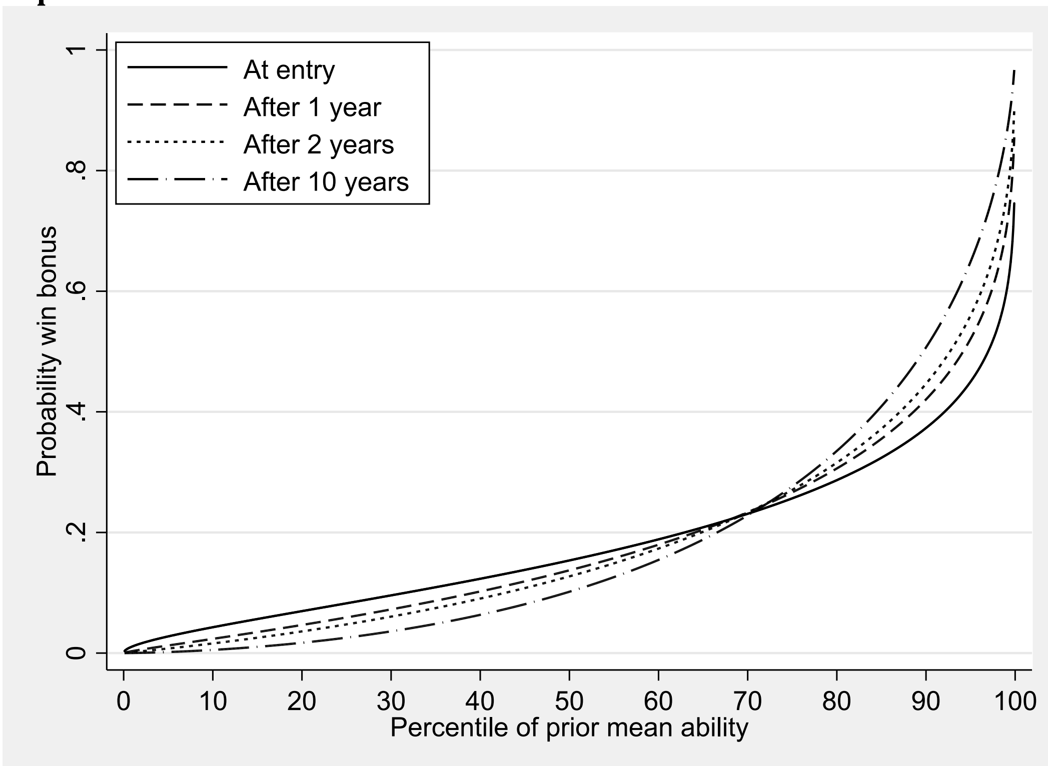

Figure 4. Average perceived probability of bonus receipt, by percentile of true ability and years of experience

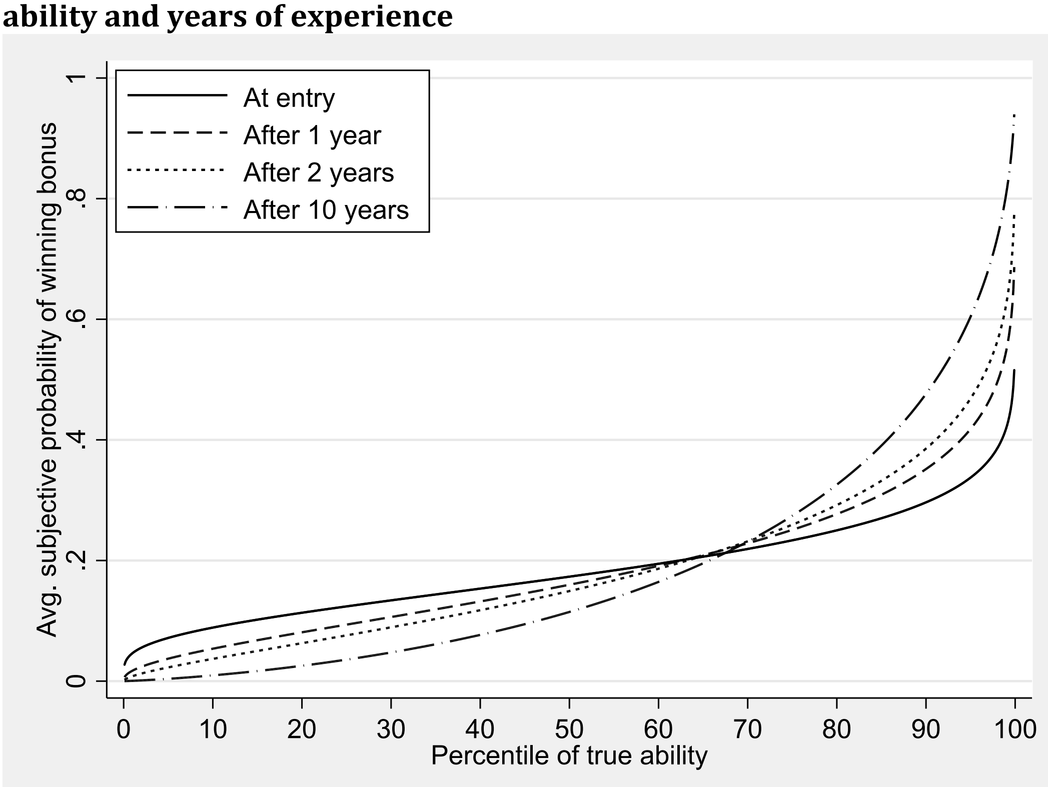

Figure 5. Effect of performance pay contract on value function, by prior percentile and years of experience

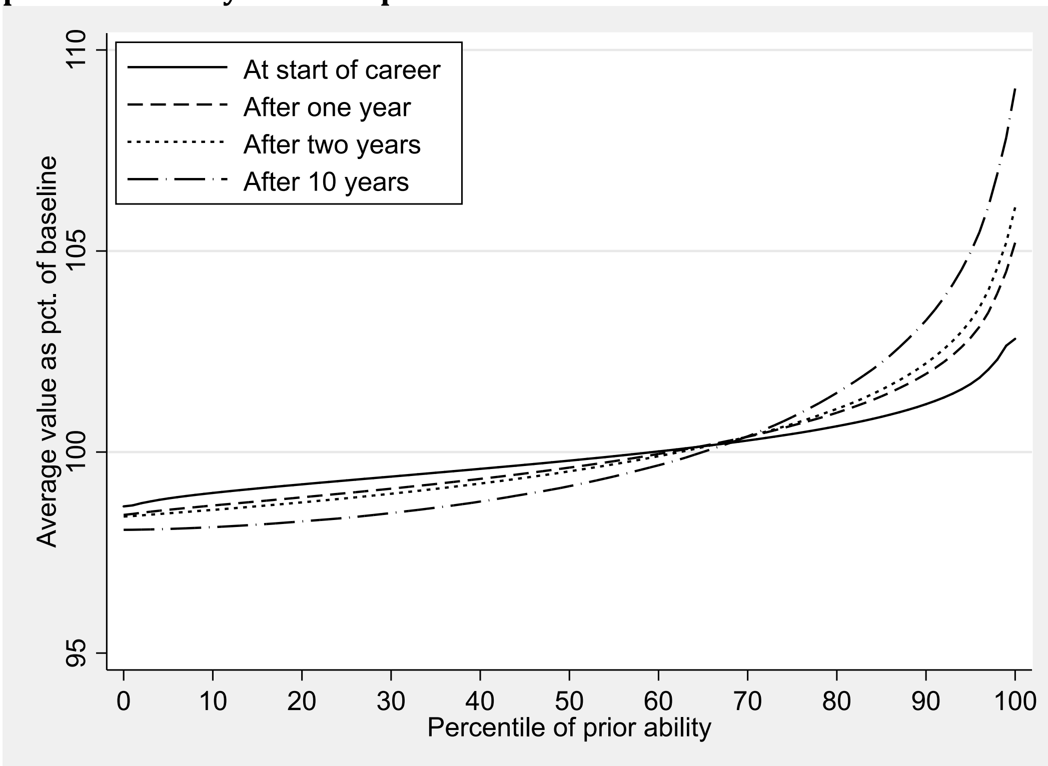

Figure 6. Effect of performance pay contract on average value function, by percentile of true ability and years of experience

Figure 7. Probability of ever being fired (assuming no quits) and average subjective expectation of firing probability at start of career, by true ability percentile

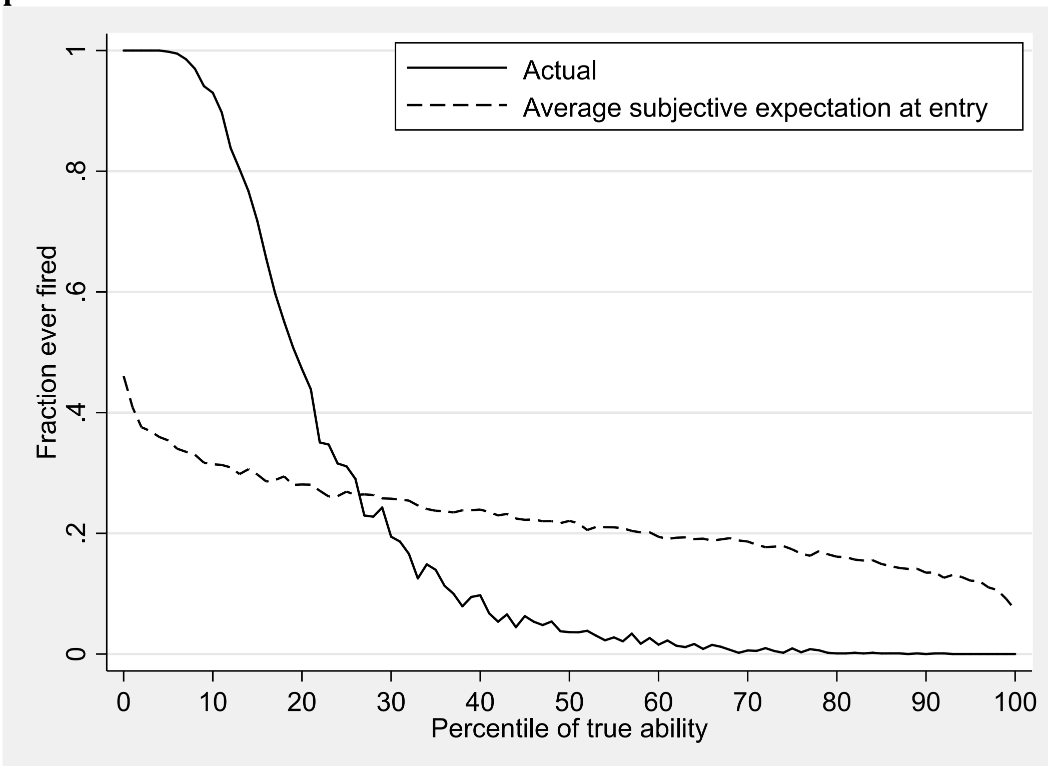

Figure 8. Effect of firing contract on average value function of teachers not yet fired, by percentile of true ability and years of experience

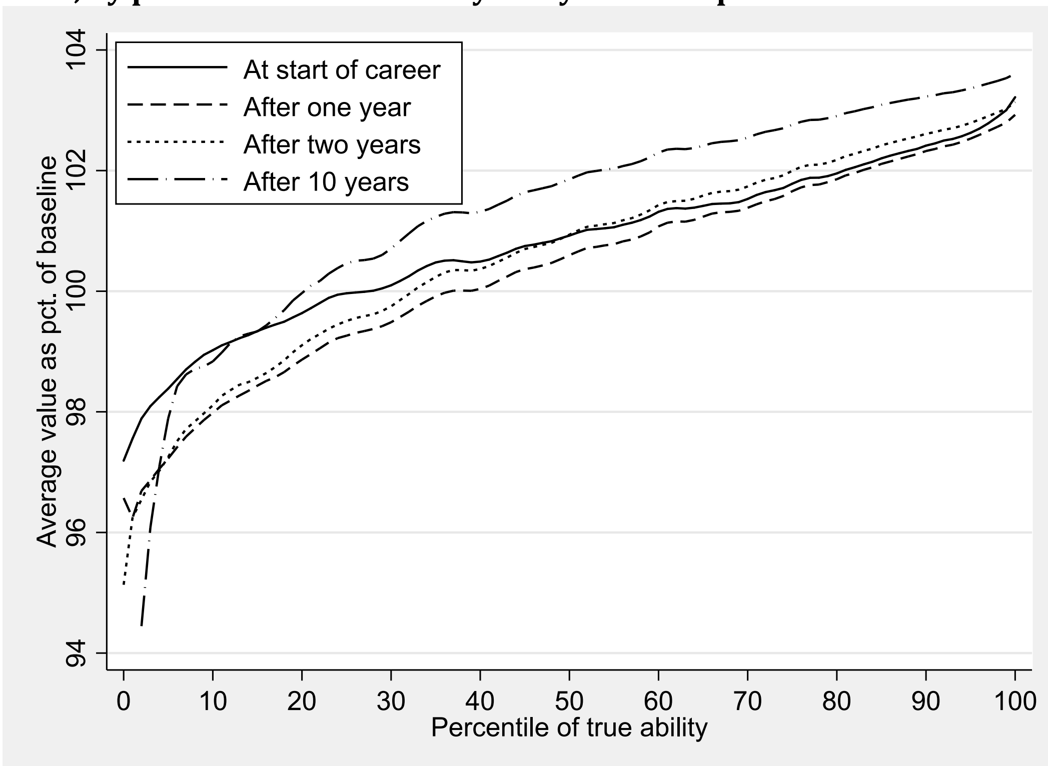

Figure 9. Effect of alternative contracts on number of new entrants to teaching, by ability percentile

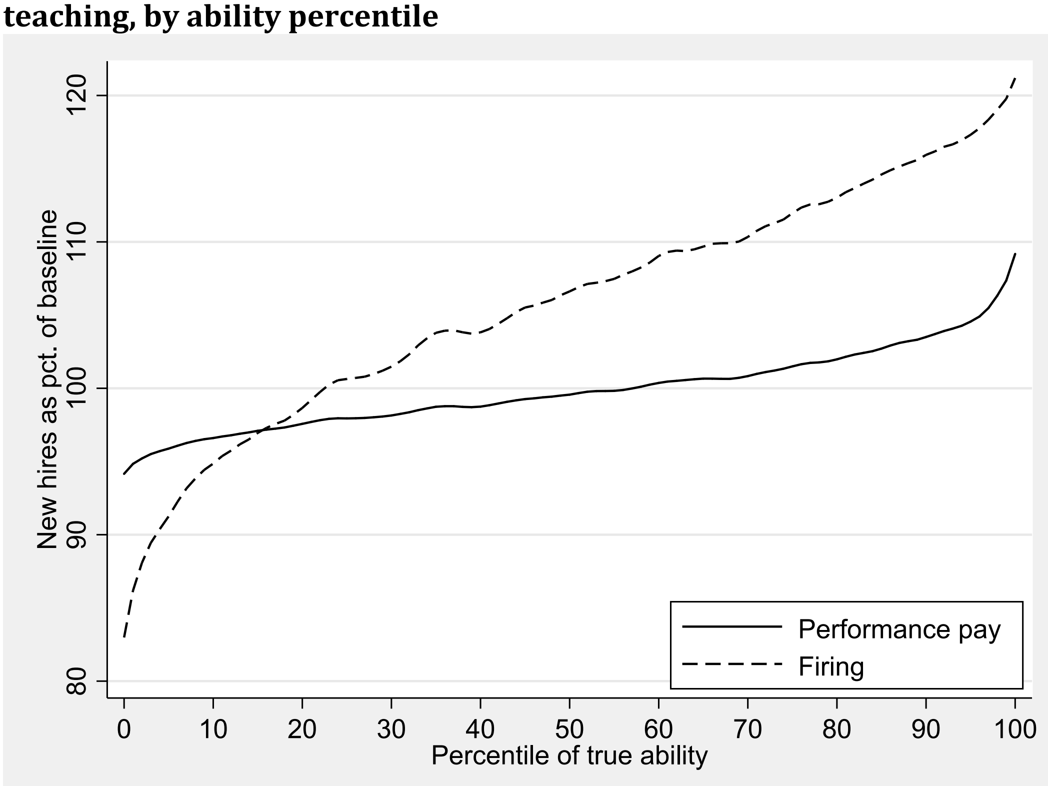

Figure 10. Effect of alternative contracts on average teaching career length, by ability percentile

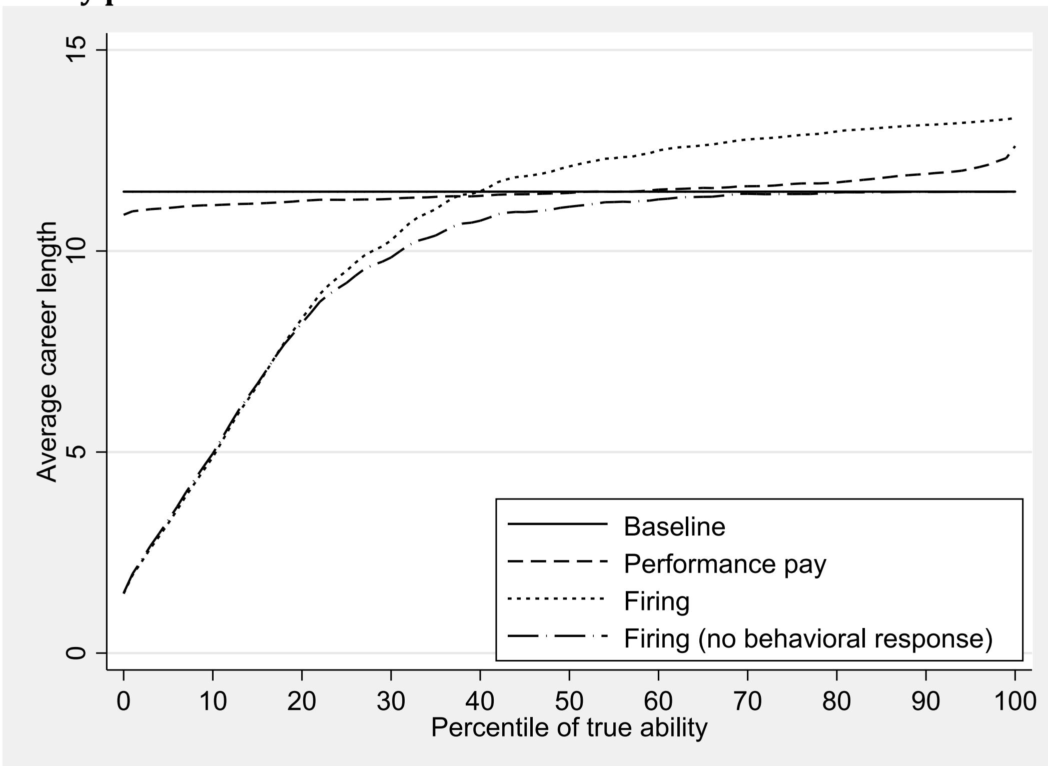

Figure 11. Effect of alternative contracts on total number of teachers, by ability percentile

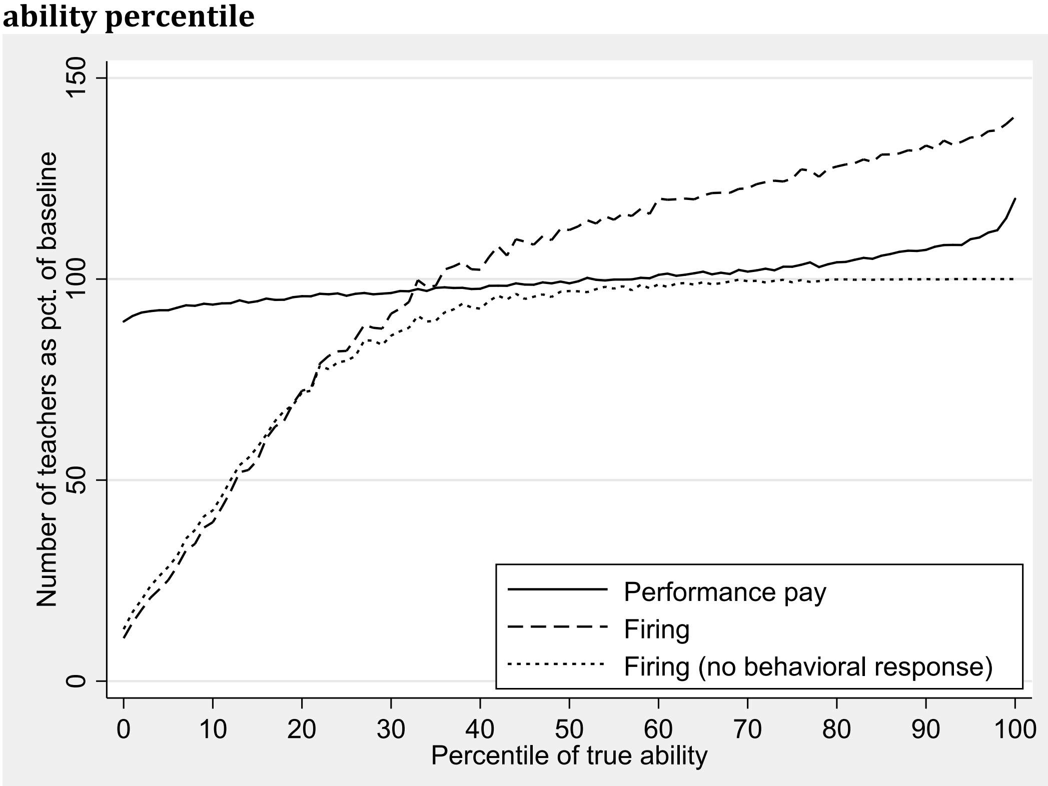

Figure 12. Effect of varying the firing rate on student achievement and total costs, baseline parameters

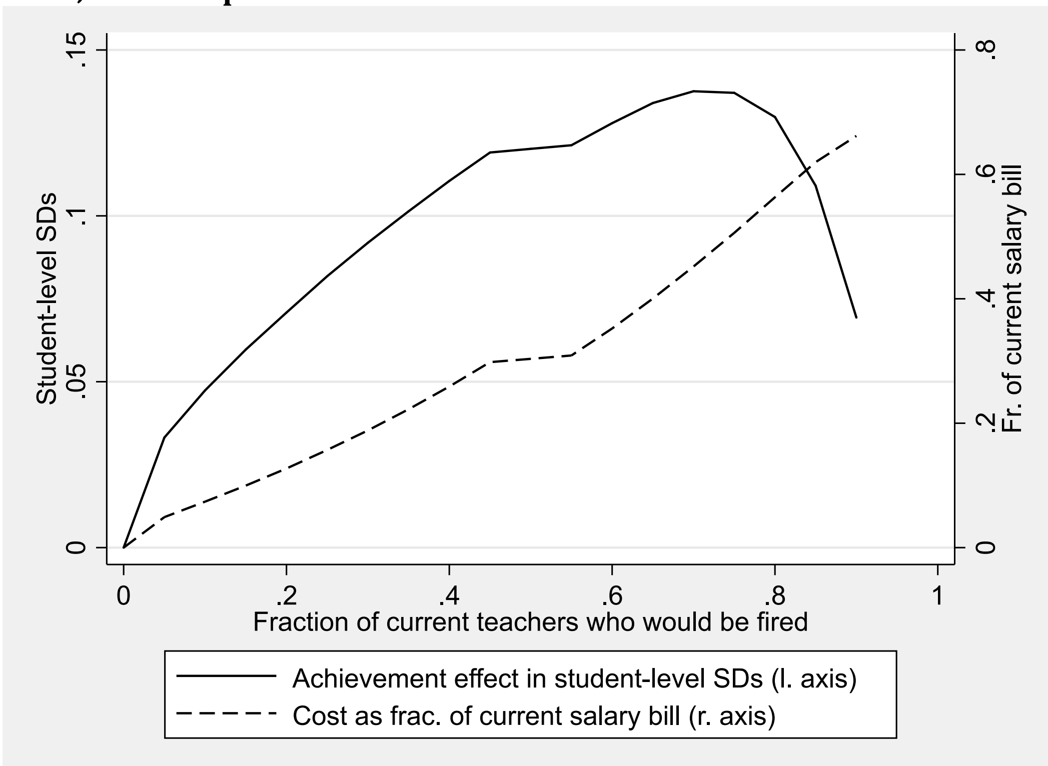

Figure 13. Effect of firing on student achievement and total costs, alternative parameters

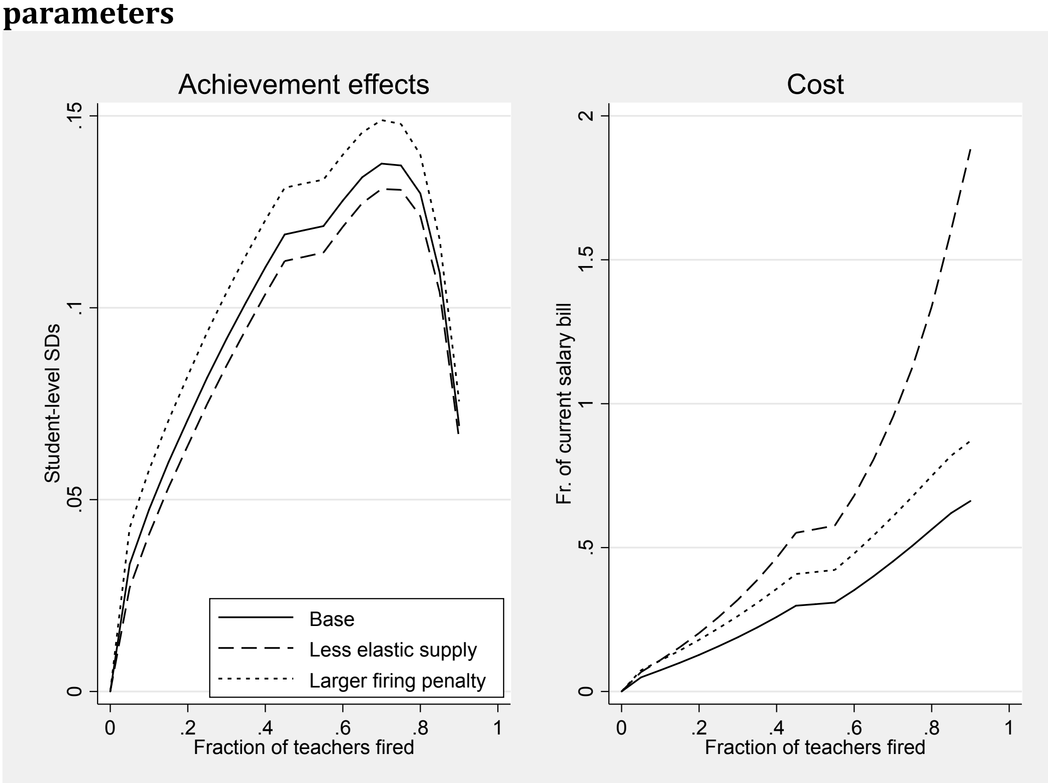

Table 2. Impact of performance pay and firing contracts on teacher effectiveness and total costs

	Baseline	Performance Pay		Firing	
		Level	Change from baseline	Level	Change from baseline
	(1)	(2)	(3)	(4)	(5)
Teacher ability (τ)					
Mean	0.000	0.007	+0.008	0.047	+0.047
SD	[0.150]	[0.150]	+0.001	[0.130]	-0.020
Teacher experience					
Mean	8.82	8.84	+0.019	9.08	+0.259
Pct. 1st year	8.0%	8.0%	-0.0 p.p.	8.2%	+0.2 p.p.
Pct. 1st 3 years	30.9%	30.8%	-0.1 p.p.	31.1%	+0.3 p.p.
Teacher effect (τ+r(t))					
Mean	-0.011	-0.003	+0.008	0.035	+0.047
SD	[0.151]	[0.152]	+0.001	[0.135]	-0.017
Salaries (expressed as multiple of baseline starting salary)					
Base starting salary	1.000	0.968	-3.2%	1.068	+6.8%
Average total pay	1.148	1.169	+1.8%	1.231	+7.3%

Table 3. Sensitivity of results to model parameters

	Performance pay		Firing	
	Effect on achievement (SDs)	Effect on salary bill (%)	Effect on achievement (SDs)	Effect on salary bill (%)
	(1)	(2)	(3)	(4)
Baseline	+0.008	+1.9%	+0.047	+7.4%
More private information				
$h=0.4$	+0.012	+1.9%	+0.050	+7.1%
$h=0.57$	+0.018	+2.0%	+0.052	+6.7%
Less noisy performance measure				
$\sigma\varepsilon=0.12$	+0.010	+2.1%	+0.048	+7.4%
Varying the supply elasticities				
$\eta=3$, $\zeta=0$	+0.004	+1.8%	+0.039	+9.2%
$\eta=1.5$, $\zeta=1.5$	+0.004	+1.8%	+0.041	+10.9%
$\eta=3$, $\zeta=1.5$	+0.006	+1.9%	+0.044	+8.2%
$\eta=1.5$, $\zeta=3$	+0.006	+1.8%	+0.044	+9.0%
$\eta=6$, $\zeta=3$	+0.012	+2.0%	+0.053	+5.9%
$\eta=3$, $\zeta=6$	+0.012	+2.0%	+0.052	+6.5%
$\eta=6$, $\zeta=6$	+0.016	+2.1%	+0.057	+5.5%
Varying the contracts				
50% bonus to 10% of teachers	+0.004	+1.9%		
20% of current teachers fired			+0.071	+12.7%

Table 4. Multi-dimensional output and influence activities that distort the performance measure

	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 reduces unmeasured effectiveness?	**n**	**n**	**n**	**y**
Impact of firing policy on:				
Effectiveness on measured dimension				
As measured	+0.047	+0.047	+0.050	+0.050
Net of influence	n/a	n/a	+0.041	+0.041
Effectiveness on unmeasured dimension	+0.047	+0.018	+0.016	+0.007
Average total pay	+7.3%	+7.3%	+6.1%	+6.1%

Appendix Figure 1. Empirical one-year attrition hazards from the Schools and Staffing Survey/Teacher Follow-up Survey