A Proofs

A.1 Main Proofs

Theorem 1. *For any weakly increasing screening function, we get that for every θ:*

$$x_1^*(\theta|p) \geq x_1^*(\theta)$$
$$x_2^*(\theta|p) \leq x_2^*(\theta).$$

Proof. As we do throughout this paper, we define $\tilde{u}(x_1, \theta)$ as the individual θ's utility when producing x_1 and then optimizing over x_2. Formally, we get that:

$$\tilde{u}(x_1, \theta) \equiv \begin{cases} \max_{e \in \mathcal{E}} u(e, \theta) & \text{if } \exists e_1 \in \mathcal{E} \text{ s.t. } e_1\theta_1 \geq x_1 \\ -\infty & \text{if } \forall e_1 \in \mathcal{E} \quad e_1\theta_1 < x_1 \end{cases} \tag{25}$$

With this definition and that $x_1^*(\theta)$ is an optimum, we get that $\tilde{u}(x_1^*(\theta), \theta) \geq \tilde{u}(x_1, \theta)$ for all $x_1 < x_1^*(\theta)$. It is then also clear that $\tilde{u}(x_1^*(\theta), \theta) + \lambda p(x_1^*(\theta))\Delta v(\theta) \geq \tilde{u}(x_1, \theta) + \lambda p(x_1)\Delta v(\theta)$ for all $x_1 < x_1^*(\theta)$ and so $x_1^*(\theta|p) \geq x_1^*(\theta)$.

Next, we define $x_2^*(x_1, \theta)$ to be θ's optimal choice x_2 when producing x_1. Importantly, this function is the same regardless of whether the utility includes the term $\lambda p(x_1)\Delta v(\theta)$ or not, since that expression does not depend on x_2. In addition, our assumptions on the cross derivatives and the assumption that $\mathcal{E}$ is convex gives us that $x_2^*(x_1, \theta)$ is decreasing in x_1. Since $x_1^*(\theta|p) \geq x_1^*(\theta)$ it therefore follows that $x_2^*(\theta|p) \leq x_2^*(\theta)$.

□

Lemma 1. *Consider two individuals θ and θ' with $x_1^*(\theta) = x_1^*(\theta')$ and $x_2^*(\theta) < x_2^*(\theta')$ and define:*

$$e^*(\theta|f) = \arg\max_{e \in \mathcal{E}} u(e, \theta) + \lambda f(x_1) \tag{4}$$
$$x_k^*(\theta|f) = e_k^*(\theta|f) \cdot \theta_k \tag{5}$$

Then $x_1^*(\theta|f) \leq x_1^*(\theta'|f)$ for any weakly increasing function $f(x_1)$ if either of the following are true:

- *λ is sufficiently large;*
- *$\tilde{u}'(x_1, \theta') - \tilde{u}'(x_1, \theta)$ is increasing in x_1*

Proof. To show that the theorem holds for sufficiently large λ, let e_1^{max} be the maximum possible effort level on task one, while still satisfying the constraint that $e \in \mathcal{E}$. Then there exists $\overline{\lambda}$ such that $\forall\lambda \geq \overline{\lambda}$, we have $\tilde{u}(\theta e_1^{max}, \theta') + \lambda f(\theta e_1^{max}) > \tilde{u}(x_1, \theta') + \lambda f(x_1)$ for every x_1 such that $f(x_1) < f(\theta e_1^{max})$. Note that this stems from the fact that $\theta_1' > \theta_1$ and so it will require less effort (on task one) to produce $x_1 = \theta e_1^{max}$ for individual θ' than individual θ. This, along with our assumption that $b(x)$ is bounded and that $c(e) < \infty$ for every $e \in \mathcal{E}$ is what allows us to conclude that the inequality holds. Of course, the inequality implies that individual θ' will prefer θe_1^{max} to any x_1 with a lower value of f and we also know that $x_1^*(\theta|f) \leq \theta e_1^{max}$ from the constraint set $\mathcal{E}$. That result is sufficient if f is strictly increasing, but since it is only weakly increasing we need to add a few more technical details. Specifically, let $\tilde{x}_1 = min\{x_1|f(x_1) = f(\theta e_1^{max})\}$, which is well-defined if f is right-continuous. From the concavity of u and the reasons discussed above, we can then conclude that $x_1^*(\theta|f) \leq max\{\tilde{x}_1, x_1^*(\theta)\} \leq x_1^*(\theta'|f)$.

See Appendix B for a proof that the theorem holds if $\tilde{u}(x_1, \theta') - \tilde{u}(x_1, \theta)$ is increasing in x_1 and a more detailed discussion of that condition. □

Theorem 2. *Consider $\theta < \theta'$ with $x_1(\theta) = x_1(\theta')$. Assume that $\Delta v(\theta)$ is increasing in θ and that either $\tilde{u}'(x_1, \theta') - \tilde{u}'(x_1, \theta)$ is increasing in x_1 or λ is sufficiently large. Then $x_1^*(\theta|p) \leq x_1^*(\theta'|p)$ for any weakly increasing screening function $p(x_1)$.*

Proof. Define $\tilde{x}_1^*(\theta'|p) = \arg\max u(x, \theta') + \lambda p(x_1) \cdot \Delta v(\theta)$, i.e., the optimal choice of individual θ' if her value of staying in the profession relative to the outside option were $\Delta v(\theta)$ instead of $\Delta v(\theta')$. From the fact that $\Delta v(\theta)$ is increasing in θ and $\theta < \theta'$, it follows that $x_1^*(\theta'|p) \geq \tilde{x}_1^*(\theta'|p)$. But from the previous theorem and the assumption that $\tilde{u}'(x_1, \theta') - \tilde{u}'(x_1, \theta)$ is increasing in x_1, we get that $\tilde{x}_1^*(\theta'|p) \geq x_1^*(\theta|p)$. Thus, $x_1^*(\theta|p) \leq x_1^*(\theta'|p)$. □

Theorem 3. *Assume the conditions on $V(\theta)$ specified above and that the assumptions in Theorem 2 hold. Furthermore, assume that θ is continuously distributed. Then for any ex post screening policy $p(x_1)$, there is a screening policy $\tilde{p}(x_1)$ that is more efficient than $p(x_1)$:*

$$\mathbb{E}_\Theta\Big[\tilde{p}\big(x_1^*(\theta|\tilde{p})\big)\Big] = \mathbb{E}_\Theta\Big[p\big(x_1^*(\theta)\big)\Big]$$
$$\mathbb{E}_\Theta\Big[V(\theta) \cdot \tilde{p}\big(x_1^*(\theta|\tilde{p})\big)\Big] \geq \mathbb{E}_\Theta\Big[V(\theta) \cdot p\big(x_1^*(\theta)\big)\Big].$$

Proof. Start with any ex post screening policy. Since $\mathbb{E}[V(\theta)|x_1(\theta)]$ is increasing in $x_1(\theta)$, it follows that an ex post screening policy with a threshold screening function, i.e., $p(x_1) = \mathbf{1}(x_1 \geq \overline{x}_1)$ for some $\overline{x}_1$, is more efficient at screening than the initial ex post screening policy. Next, choose $\tilde{x}_1$ such that the announced screening policy with a threshold screening function $\tilde{p}(x_1) = \mathbf{1}(x_1 \geq \tilde{x}_1)$ retains the same fraction of teachers, i.e. $\mathbb{E}_\Theta\Big[\tilde{p}\big(x_1^*(\theta|\tilde{p})\big)\Big] = \mathbb{E}_\Theta\Big[p\big(x_1^*(\theta)\big)\Big]$. We show in Lemma A.4 that under the assumption that

$x_1^*(\theta)$ is continuously distributed such an $\tilde{x}$ exists. Using Lemma A.2, we can conclude that $\mathbb{E}_\Theta\Big[V(\theta)\cdot\tilde{p}\big(x_1^*(\theta|\tilde{p})\big)\Big] \geq \mathbb{E}_\Theta\Big[V(\theta)\cdot p\big(x_1^*(\theta)\big)\Big]$, which by definition means that the announced screening policy is more efficient at screening than the ex post screening policy. Since we did not specify the initial ex post screening policy the result that there is always an announced screening policy that is more efficient at screening is true for all ex post screening policy policies. □

A.2 Supporting Lemmas

Lemma A.1. *Suppose that $x_1^*(\theta) = x_1^*(\theta')$ and $x_2^*(\theta') > x_2^*(\theta)$. Then $\theta'_k > \theta_k$ for $k \in \{1,2\}$.*

Proof. First, we note that $x_1^*(\theta)$ is increasing in θ_1 and decreasing in θ_2. Thus, if $\theta'_1 \geq \theta_1$ and $\theta'_2 \leq \theta_2$, then by our assumptions we would have that $x_1^*(\theta') > x_1^*(\theta)$. (This assumes that the comparative statics are strict.). We can similarly rule out the fact that $\theta'_1 \leq \theta_1$ and $\theta'_2 \geq \theta_2$. We can also rule out the fact that $\theta'_2 < \theta_2$. (Otherwise, it would be then be cheaper for θ than θ' to move from $x_2^*(\theta)$ to $x_2^*(\theta')$, which would contradict the assumption of optimization.) This rules out any other case than $\theta'_k > \theta_k$ for $k \in \{1,2\}$. □

Lemma A.2. *Assume the conditions in Theorem 2 and the assumptions on $V(\theta)$ outlined in the paper. Then consider a ex post screening policy with a threshold screening function $p(x_1) = \mathbf{1}(x_1 \geq \overline{x_1})$ and an announced screening policy with a threshold screening function $p(x_1) = \mathbf{1}(x_1 \geq \tilde{x}_1)$. Define:*

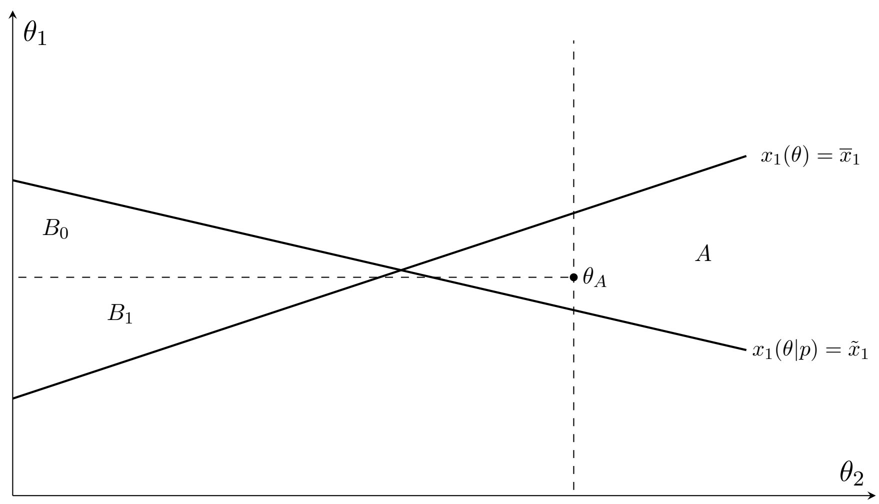
$$A = \{\theta | x_1(\theta) < \overline{x}_1 \ \&\ x_1(\theta|p) \geq \tilde{x}_1\}$$
$$B = \{\theta | x_1(\theta) \geq \overline{x}_1 \ \&\ x_1(\theta|p) < \tilde{x}_1\}$$

Then $\forall \theta^A \in A, \theta^B \in B$, we have that $V(\theta^A) \geq V(\theta^B)$.

Proof Sketch. To illustrate the intuition, we plot the parameter space (θ_1, θ_2) below. On this space we draw a line through all of the points in which $x_1(\theta) = \overline{x_1}$. From our assumptions, every θ point below this line has $x_1(\theta) < \overline{x_1}$ and so is not retained under the ex post screening policy, while every θ above the line has $x_1(\theta) \geq \overline{x_1}$ and so is retained. While we could similarly label the line $x_1(\theta|p) \geq \tilde{x}_1$ in the announced policy, there will likely be bunching there and so the line should more precisely be defined as partitioning the space into the θ's that are retained under the announced screening policy and those that are not.

From the results of Theorem 2, we get that the slope of the line where $x_1(\theta|p) = \tilde{x}_1$ is less positive than the line where $x_1(\theta) = \overline{x}_1$, as they are illustrated below. From this, we can conclude that $\forall \theta^A \in A, \theta^B \in B$, we have that $\theta_2^A > \theta_2^B$. If $\theta_B \in B_1$ in the figure below, i.e., that $\theta_1^A > \theta_1^B$, then $\theta_A > \theta_B$ and so it follows directly that $V(\theta_A) > V(\theta_B)$. The challenging case is therefore when $\theta_1^B > \theta_1^A$, i.e., when $\theta_B \in B_0$.

Under the assumption that $\theta_1^B > \theta_1^A$ and $\theta_2^A > \theta_2^B$, however, it follows that moving from $x_1(\theta)$ to $\tilde{x}_1$ is "cheaper" for θ_B than for θ_A in the sense that $\tilde{u}(\tilde{x}_1, \theta_B) - u(x_1(\theta_B), \theta_B) \geq \tilde{u}(\tilde{x}_1, \theta_A) - u(x_1(\theta_A), \theta_A)$. This, along with the fact that θ_A does increase to $\tilde{x}_1$ under the announced screening policy but θ_B does not, allows us to infer that $\Delta v(\theta_A) > \Delta v(\theta_B)$. Combining this result with the fact that $\theta_1^B > \theta_1^A$ and $\theta_2^A > \theta_2^B$ and the assumptions regarding V, we get that $V(\theta_A) > V(\theta_B)$.



□

Proof. Throughout, we will assume that $A, B \neq \emptyset$, since otherwise the statement is vacuous. We fix $\theta^A \in A$ and $\theta^B \in B$. We start by ruling out the possibility that $\theta^A \leq \theta^B$, i.e., that $\theta_k^A \leq \theta_k^B$ for $k \in \{1, 2\}$. If it were, then there would exist a $\tilde{\theta}$ such that $\tilde{\theta}_2 = \theta_2^B$, $\tilde{\theta}_1 < \theta_1^B$, and $x_1(\theta^A) = x_1(\tilde{\theta})$ (see Lemma A.3 for the proof). From Lemma 1, we then get that $x_1(\tilde{\theta}|p) \geq x_1(\theta^A|p)$. However, since $\tilde{\theta}_2 = \theta_2^B$ and $\tilde{\theta}_1 < \theta_1^B$, we also get that $x_1(\theta^B|p) \geq x_1(\tilde{\theta}|p)$ since $x_1(\theta|p)$ is increasing in θ_1. Therefore, $x_1(\theta^B|p) \geq x_1(\theta^A|p)$, which is a contradiction to the fact that $\theta^A \in A$ and $\theta^B \in B$.

Similarly, it also cannot be the case that $\theta_2^B \geq \theta_2^A$ and $\theta_1^B \leq \theta_1^A$. This follows from the fact that $x_1(\theta)$ is increasing in θ_1 and decreasing in θ_2 and $x_1(\theta^B) > x_1(\theta^A)$. Thus, we can infer that $\theta_2^B \leq \theta_2^A$.

If $\theta_1^B \leq \theta_1^A$ then $\theta^A \geq \theta^B$and so by assumption $V(\theta^A) \geq V(\theta^B)$. Thus, in what follows we will assume that $\theta_1^B \geq \theta_1^A$ and $\theta_2^B \leq \theta_2^A$ and show that in this case $V(\theta^A) \geq V(\theta^B)$. To do so, we define $\tilde{u}(x_1, \theta)$ as in the paper, i.e.,

$$\tilde{u}(x_1, \theta) \equiv \begin{cases} \max_{e \in \mathcal{E}} u(e, \theta) & \text{if } \exists e_1 \in \mathcal{E} \text{ s.t. } e_1\theta_1 \geq x_1 \\ -\infty & \text{if } \forall e_1 \in \mathcal{E} \quad e_1\theta_1 < x_1 \end{cases} \tag{26}$$

Since $\theta^A \in A$ and $\theta^B \in B$, we know that θ^A finds it worth producing $\tilde{x}_1$ in the pre-tenure period under the announced screening policy, while θ^B does not. From this, we can conclude that:

$$v(\theta^A) \leq \tilde{u}(\tilde{x}_1, \theta^A) + \lambda \cdot \Delta v(\theta^A) \tag{27}$$

$$v(\theta^B) \geq \tilde{u}(\tilde{x}_1, \theta^B) + \lambda \cdot \Delta v(\theta^B) \tag{28}$$

Rearranging, we get that:

$$\lambda \cdot \Big[\Delta v(\theta^A) - \Delta v(\theta^B)\Big] \geq \Big[\big(\tilde{u}(\tilde{x}_1, \theta^B) - v(\theta^B)\big) - \big(\tilde{u}(\tilde{x}_1, \theta^A) - v(\theta^A)\big)\Big] \tag{29}$$

Finally, from the envelope theorem we get that that $\tilde{u}(x, \theta) - v(\theta)$ is increasing in θ_1 and decreasing in θ_2 for every x. Thus, $\Big[\big(\tilde{u}(\tilde{x}_1, \theta^B) - v(\theta^B)\big) - \big(\tilde{u}(\tilde{x}_1, \theta^A) - v(\theta^A)\big)\Big] \geq 0$ and so $\Delta v(\theta^A) \geq \Delta v(\theta^B)$. This, combined with the assumptions about V and the fact that $\theta_2^B \leq \theta_2^A$ and $\theta_1^B \geq \theta_1^A$ implies that $V(\theta^A) \geq V(\theta^B)$. □

Lemma A.3. *Suppose that $\theta^A \in A$ and $\theta^B \in B$ and $\theta^A \leq \theta^B$. Then exists a $\tilde{\theta}$ such that $\tilde{\theta}_2 = \theta_2^B$, $\tilde{\theta}_1 \leq \theta_1^B$, and $x_1^*(\theta^A) = x_1^*(\tilde{\theta})$.*

Proof. Since $x_1^*(\theta)$ is decreasing in θ_2 and by assumption $x_1^*(\theta^A) < x_1^*(\theta^B)$ and $\theta_2^B \leq \theta_2^A$, it follows that $x_1^*(\theta_1^A, \theta_2^B) < x_1^*(\theta^A)$. From Berge's maximization theorem, $x_1^*(\theta)$ is upper hemicontinuous and so from the intermediate value theorem as we increase θ_1 from (θ_1^A, θ_2^B) to (θ_1^B, θ_2^B) there must be a $\tilde{\theta}_1 \in (\theta_1^A, \theta_1^B)$ such that $x_1^*(\theta^A) = x_1^*(\tilde{\theta}_1, \theta_2^B)$. □

Lemma A.4. *Suppose that θ is continuously distributed. Then for every $p \in (0, 1)$ there exists a $\tilde{x}$ such that under the policy $\tilde{p}(x_1) = \mathbf{1}(x_1 \geq \tilde{x}_1)$, we get that $\mathbb{E}_\Theta\Big[\tilde{p}\big(x_1^*(\theta|\tilde{p})\big)\Big] = p$.*

Proof. Define $g(\tilde{x}_1) \equiv \mathbb{E}_\Theta\Big[\tilde{p}\big(x_1^*(\theta|\tilde{p})\big)\Big]$. If g is a continuous function of $\tilde{x}_1$, the intermediate value theorem implies the result.

To show that g is continuous, we consider a sequence that converges to $\tilde{x}_1$, i.e., $(\tilde{x}_{1,n}) \to \tilde{x}_1$. If $g(\tilde{x}_{1,n}) \to g(\tilde{x}_1)$, then it follows that g is continuous.

By assumption θ is continuously distributed, and we will denote its probability density function as $f(\theta)$. Then we can write $g(\tilde{x}_1)$ as:

$$g(\tilde{x}_1) = \int \mathbf{1}\Big(\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) \geq 0\Big) f(\theta) d\theta. \tag{30}$$

Define $h(\theta) \equiv \mathbf{1}\Big(\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) \geq 0\Big)$ and $h_n(\theta) \equiv \mathbf{1}\Big(\tilde{u}(\tilde{x}_{1,n}, \theta) + \lambda \Delta v(\theta) - v(\theta) \geq 0\Big)$. Also define $e_{1,max}$ as $sup\{e_1 | e \in \mathcal{E}\}$. From Lemma A.5, $h_n(\theta) \to h(\theta)$ at every θ such that $\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) \neq 0$ and $\theta_1 e_{1,max} \neq \tilde{x}_1$. Furthermore, from our assumption that $V(\theta)$ is increasing in θ_1 and that $\frac{\partial V(\theta)}{\partial \theta_1} \leq \frac{\partial \Delta v(\theta)}{\partial \theta_1}$ and the envelope condition, we get that $\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta)$ is strictly increasing in θ_1 at every point where $\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) = 0$. This means there is at most one θ_1 for every θ_2 such that $\tilde{u}(\tilde{x}_{1,n}, \theta) + \lambda \Delta v(\theta) - v(\theta) = 0$. Under the assumption that θ is continuously distributed, this implies that the set $\{\theta | \tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) = 0 \cup \theta_1 e_{1,max} = \tilde{x}_1\}$ has zero measure. Together with the result from Lemma A.5, this implies that $h_n \to h$ pointwise almost everywhere. From the bounded convergence theorem, it follows that $g_n \to g$ and so g is continuous. □

Lemma A.5. *Let $e_{1,max} = sup\{e_1 | e \in \mathcal{E}\}$. Then $h_n(\theta) \to h(\theta)$ at every θ such that $\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) \neq 0$ and $\tilde{x}_1 \neq \theta_1 e_{1,max}$.*

Proof. Consider some θ such that $\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) \equiv \epsilon \neq 0$. We first consider the case in which $\tilde{x}_1 < e_{1,max}\theta_1$. We then know that $\tilde{u}(\tilde{x}_1, \theta)$ is continuous in $\tilde{x}_1$ and so there exists a $\delta > 0$ such that $|\tilde{u}(\tilde{x}_1, \theta) - \tilde{u}(x_1, \theta)| < \frac{|\epsilon|}{2}$ for every x_1 such that $|x_1 - \tilde{x}_1| < \delta$. This implies that $\mathbf{1}\Big(\tilde{u}(x_1, \theta) + \lambda \Delta v(\theta) - v(\theta) \geq 0\Big) = \mathbf{1}\Big(\tilde{u}(\tilde{x}_1, \theta) + \lambda \Delta v(\theta) - v(\theta) \geq 0\Big)$ for every x_1 such that $|x_1 - \tilde{x}_1| < \delta$. It thus follows that there exists an N such that $h_n(\theta) = h(\theta)$ for all $n > N$.[46]

If $\tilde{x}_1 > e_{1,max}\theta_1$, then $\tilde{u}(\tilde{x}_1, \theta) = -\infty$. Again, there exists an N such that $h_n(\theta) = h(\theta)$ for all $n > N$ since for all $n > N$ we get that $\tilde{x}_{1,n} > e_{1,max}\theta_1$.

□

B Model Appendix

B.1 The Inefficiency of Screening on Output

Much of the intuition for our theoretical results can be seen in a comparison between two infeasible screening regimes. One – which we refer to as "ex post screening" – screens on the teachers' output without the additional incentive of the screening policy in place, i.e., on $x_1^*(\theta)$ and the other – which we refer to as "ability screening" – screens on the first dimension of ability, i.e., on θ_1. Since the principals need not worry about incentives in either regime, it is easy to see that if u is differentiable and the constraint that $e \in \mathcal{E}$ does not bind, then the two regimes would be equivalent if the principal observed both dimensions of ability and both dimensions of output. There is an underlying inefficiency with screening on output, however, that appears when only one of the two outputs is observed.

This inefficiency can be summarized with the following lemma.

[46]This is because $\tilde{x}_{1,n} \to \tilde{x}_1$ means that there exists some N such that $|\tilde{x}_{1,n} - \tilde{x}_1| < \delta$ for all $n > N$.

Lemma B.1. *In any ex post screening policy, the probability that individual θ is retained is weakly decreasing in θ_2 for every θ_1.*

Proof. This follows directly from the definition of an ex post screening policy and the fact that $x_1^*(\theta)$ is decreasing in θ_2. □

In some sense, we can think of this lemma as being roughly akin to a screening version of the traditional multitasking problem. The traditional multitasking problem highlights an inherent inefficiency when the principal can only add an incentive to a single output: it reduces effort related to the production of the other outcomes. The screening version outlined in the lemma above instead highlights an inherent inefficiency when the principal can only screen on a single output: an individual who is effective at increasing the other output is less likely to be retained than one who is ineffective at increasing the other output and equally effective at increasing the output screened on. This implies the following theorem:

Thm B.1. *The ability screening regime is more efficient than the ex post screening regime.*

Proof. Consider any ex post screening policy $p(x_1)$ and define an ability screening regime as:

$$\tilde{p}(\theta_1) = \mathbb{E}[p(x_1^*(\theta)|\theta_1] \tag{31}$$

Because p and $\tilde{p}$ retain the same fraction of teachers of each θ_1, they retain the same fraction of all teachers. Furthermore, since $p(x_1^*(\theta))$ is decreasing in θ_2, for any increasing function $V(\theta)$, we have that: $\mathbb{E}[V(\theta)\tilde{p}(\theta_1)|\theta_1] \geq \mathbb{E}[V(\theta)p(x_1^*(\theta))|\theta_1]$ for every θ_1. Since that inequality holds for every θ_1, we get that: $\mathbb{E}[V(\theta)\tilde{p}(\theta_1)] \geq \mathbb{E}[V(\theta)p(x_1^*(\theta))]$. □

Again, the intuition is straightforward. Principals would like to keep teachers with both high θ_1 and θ_2 and Lemma B.1 shows that an ex post screening policy that retains individuals with high $x_1^*(\theta)$ biases one toward removing teachers with high θ_2. While not actively retaining individuals with high θ_2 the way an optimal policy would, screening on ability rather than output does at least remove this bias and is thus more efficient.

Finally, note that theorem does not say that ex post screening on $x_1^*(\theta)$ would necessarily reduce the average of $x_2^*(\theta)$ in the population. If θ_1 and θ_2 are highly correlated, ex post screening on $x_1^*(\theta)$ would improve the average $x_2^*(\theta)$ in the population; however, Theorem B.1 says that $x_2^*(\theta)$ could be increased even more if the policy maker were able to screen directly on θ_1 rather than on $x_1^*(\theta)$.

B.2 Different Responses to the Same Incentive

Suppose there is a manager who cares about two dimensions of employee output, but who can only observe one of those dimensions. In this principal-agent problem, often referred to as the multitasking problem (Holmstrom and Milgrom, 1991), the manager may shy away from implementing as large an incentive on the output she observes if doing so causes employees to shift away from time spent producing other dimension of output. But suppose that she also must evaluate the employees based on the single observed dimension at the end of the period; does that change her decision about the optimal size of incentive? Stated differently, does the substitution pattern induced by the added incentive cause the first dimension to be a more or less noisy signal of the employees' underlying ability?

As we discuss in the paper, the answer depends on how different individuals respond differently to the incentive. Here, we expand on the results in the paper to show that a very similar condition on the utility function discussed in Lemma 1 is not only sufficient, but – in a limited sense – necessary to ensure that higher ability individuals respond more to the incentive than lower ability individuals who produce the same x_1 without the incentive. We then discuss in a bit more detail the economic meaning behind the condition and provide a few examples of specific utility functions that do (and do not) meet the criteria.

B.2.1 Formal Model and Definitions:

We use the same model of employee output as described in Section II.B and consider the case where each employee gets paid – or has to pay – an additional $f(x_1)$ when producing x_1. For notational simplicity, we drop the λ term in front of f in this section; doing so in Section II.B allowed us to formally consider the case where the incentive got large, but that is not the focus here and so we will stop the over-parameterization of the f function. If $f(x_1)$ is increasing, then this serves as an additional positive incentive; if it is decreasing it serves as an additional negative incentive. A key point here is that all individuals receive the same incentive and value it equivalently in utility terms.

As in the paper, it helps to define $\tilde{u}(x_1, \theta)$ as the optimal utility individual θ can get when constrained to produce at least x_1, i.e.,

$$\tilde{u}(x_1, \theta) \equiv \begin{cases} \max_{e \in \mathcal{E}} u(e, \theta) & \text{if } \exists e_1 \in \mathcal{E} \text{ s.t. } e_1\theta_1 \geq x_1 \\ -\infty & \text{if } \forall e_1 \in \mathcal{E} \quad e_1\theta_1 < x_1 \end{cases} \tag{32}$$

We will use $\tilde{u}'(x_1, \theta)$ to denote $\frac{\partial \tilde{u}(x_1, \theta)}{\partial x_1}$ when $\tilde{u}$ is differentiable and while differentiability is not an necessary assumption in our results, it makes the notation and intuition more straightforward. Throughout this discussion, we will also focus on cases where f is small enough that we do not have to worry about $\tilde{u}(x_1, \theta)$ equalling $-\infty$ for either of the two individuals we consider, which simplifies the exposition.

We next turn to conditions on $\tilde{u}$ that allow us to make conclusions about which individuals respond more or less to the added incentive. The condition, which is both sufficient and – in a limited sense we discuss below – necessary, is a single crossing condition on the derivative of $\tilde{u}$, defined formally below.

Definition 1. *We say that $\tilde{u}(x_1, \theta)$ has the* single crossing condition on $\tilde{u}'$ *if every $\theta' > \theta$ with $x_1^*(\theta) = x_1^*(\theta') \equiv x_1^*$, $\tilde{u}(x_1, \theta') - \tilde{u}(x_1, \theta)$ is strictly decreasing for all $x_1 < x_1^*$ and $\tilde{u}(x_1, \theta') - \tilde{u}(x_1, \theta)$ is strictly increasing for all $x_1 > x_1^*$.*

We call this a single crossing condition because – in the case where $\tilde{u}$ is differentiable – it implies that $\tilde{u}'(x_1, \theta)$ and $\tilde{u}'(x_1, \theta')$ cross a single time, at x_1^*. We state this formally in the following remark:

Remark 1. *When $\tilde{u}(x_1, \theta)$ is differentiable, the two definitions are equivalent:*

1. *$\tilde{u}(x_1, \theta)$ has the single crossing condition on $\tilde{u}'$ as defined above;*

2. *For every $\theta' > \theta$ with $x_1^*(\theta) = x_1^*(\theta') \equiv x_1^*$, we have that $\tilde{u}'(x_1, \theta') > \tilde{u}'(x_1, \theta)$ for every $x_1 > x_1^*$ and $\tilde{u}'(x_1, \theta') < \tilde{u}'(x_1, \theta)$ for every $x_1 < x_1^*$.*

We deliberately choose to call this a single crossing condition on $\tilde{u}'$ to highlight the the similarities and differences between our results and the traditional results on comparative statics. While traditional comparatives statics aims at understanding how the optimal choice of x varies according to characteristics θ, we consider a slightly different question and aim to understand how the *change* in the optimal choice of x vary according to characteristics θ when the individuals are presented with an identical *change* in incentives. Interestingly, the result mirrors the result from comparative statics, although it now hinges on increasing differences in the *marginal*, i.e., change in, utility function rather than the utility function itself. Just as in traditional comparative statics, we can also relax our condition slightly from a single-crossing condition to an increasing differences condition, which is no longer necessary but is sufficient; again, this increasing differences condition is on the marginal utility rather than the utility function itself and can be stated as a condition on the second derivatives of $\tilde{u}$. We state this formally in the remark below.

Remark 2. *If $\tilde{u}''(x_1, \theta') > \tilde{u}''(x_1, \theta)$ for every $\theta' > \theta$ with $x_1^*(\theta) = x_1^*(\theta') \equiv x_1^*$, then $\tilde{u}'(x_1, \theta)$ has the single crossing condition on $\tilde{u}'$ as defined above.*

A key difference between our results and traditional comparative statics results is that ours are much more limited, in that we do not compare the change in optimal choices of any two individuals but only individuals who choose the same output absent the additional incentive. This also handles the challenges inherent to settings with multidimensional types. By narrowing our comparison to individuals who choose the same output absent the additional incentive, we essentially reduce the types to a single dimension. Because we condition on a choice ($x_1(\theta)$) rather than a type, the analysis does not rely on standard single-dimensional screening results.

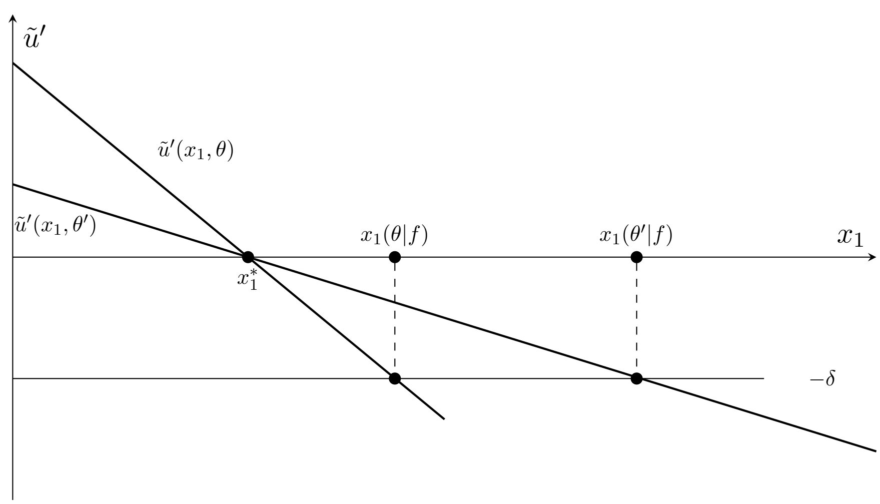
B.2.2 Theory Results and Proofs:

Given the definition of the single crossing condition on $\tilde{u}'$ from above, we can finally turn to the results. First, we present the result that the single crossing condition is sufficient to ensure that θ' responds more to the incentive than θ.

Thm B.2. *Assume that $\tilde{u}(x_1, \theta)$ has the single crossing condition on $\tilde{u}'$, and consider any $\theta' > \theta$ with $x_1^*(\theta) = x_1^*(\theta')$. Then:*

- *For every weakly increasing function $f(x_1)$, we have $x_1^*(\theta'|f) \geq x_1^*(\theta|f)$. Similarly, for every weakly decreasing function $x_1^*(\theta'|f) \leq x_1^*(\theta|f)$.*
- *Further assume that $\tilde{u}$ is differentiable. Then for every strictly increasing differentiable function $f(x_1)$, we have $x_1^*(\theta'|f) > x_1^*(\theta|f)$. Similarly, for every strictly decreasing continuous function $x_1^*(\theta'|f) < x_1^*(\theta|f)$.*

Proof Sketch. Suppose for simplicity that $f(x_1) = \delta \cdot x_1$ for some $\delta > 0$ and consider the figure below, where the curves $\tilde{u}'(x_1, \theta')$ and $\tilde{u}'(x_1, \theta)$ are both plotted on a graph with $\tilde{u}'$ on the y-axis and x_1 on the x-axis. From the assumptions on u, we get that both $\tilde{u}'(x_1, \theta')$ and $\tilde{u}'(x_1, \theta)$ are downward slopping and if we assume $\tilde{u}$ is convex, the optimal choice of $x_1^*(\theta|f)$ is the point where $\tilde{u}'(x_1, \theta) = -\delta$. Clearly $x_1^*(\theta|f) > x_1^*$ and from the single crossing condition on $\tilde{u}'$, we get that $\tilde{u}'(x_1, \theta') > \tilde{u}'(x_1, \theta)$ for all $x_1 > x_1^*$, so it follows that $\tilde{u}'(x_1^*(\theta|f), \theta') > -\delta$. Thus, $x_1^*(\theta'|f) > x_1^*(\theta|f)$.



□

Proof. Consider any weakly increasing function $f(x_1)$. Clearly, $x_1^*(\theta|f) \geq x_1^*$ and $x_1^*(\theta'|f) \geq x_1^*$. If $x_1^*(\theta|f) = x_1^*$, then we are done, so we will assume in what follows that $x_1^*(\theta|f) > x_1^*$.

We then consider any $x_1 \in (x_1^*, x_1^*(\theta|f))$. Since $x_1^*(\theta|f)$ is an optimizer, we get that $\tilde{u}(x_1,\theta)+f(x_1) \leq \tilde{u}(x_1^*(\theta|f),\theta) + f(x_1^*(\theta|f))$ or that $\tilde{u}(x_1^*(\theta|f),\theta) - \tilde{u}(x_1,\theta) \geq f(x_1) - f(x_1^*(\theta|f))$. Further, since $x_1^*(\theta|f) > x_1 \geq x_1^*$ we get from the single crossing condition on $\tilde{u}'$ that: $\tilde{u}(x_1^*(\theta|f),\theta') - \tilde{u}(x_1^*(\theta|f),\theta) > \tilde{u}(x_1,\theta') - \tilde{u}(x_1,\theta)$. Rearranging, it follows that:

$$\tilde{u}(x_1^*(\theta|f),\theta') - \tilde{u}(x_1,\theta') > \tilde{u}(x_1^*(\theta|f),\theta) - \tilde{u}(x_1,\theta) \geq f(x_1) - f(x_1^*(\theta|f)).$$

Thus, individual θ' would choose $x_1^*(\theta|f)$ over x_1 for all $x_1 \in (x_1^*, x_1^*(\theta|f))$, which along with the fact that $x_1^*(\theta'|f) \geq x_1^*$, proves that $x_1^*(\theta'|f) \geq x_1^*(\theta|f)$. The proof that $x_1^*(\theta'|f) \leq x_1^*(\theta|f)$ for any weakly decreasing function f is identical.

To prove the second bullet point, that if $\tilde{u}$ is differentiable and $f(x_1)$ is a strictly increasing differentiable function we get that $x_1^*(\theta'|f) > x_1^*(\theta|f)$, we note that the single crossing condition implies: $\tilde{u}'(x_1^*(\theta|f),\theta') > \tilde{u}'(x_1^*(\theta|f),\theta)$. Furthermore, since $x_1^*(\theta|f)$ is an optimum, we get that $\tilde{u}'(x_1^*(\theta|f),\theta) = -f'(x_1^*(\theta|f))$ for interior $x_1^*(\theta|f)$. Together, this implies that $\tilde{u}'(x_1^*(\theta|f),\theta') + f'(x_1^*(\theta|f)) > 0$. Thus, for small enough $\epsilon > 0$, we get that $\tilde{u}(x_1^*(\theta|f)+\epsilon,\theta') + f(x_1^*(\theta|f)+\epsilon) > \tilde{u}(x_1^*(\theta|f),\theta') + f(x_1^*(\theta|f))$ and so θ' would choose $x_1^*(\theta|f)+\epsilon$ over $x_1^*(\theta|f)$. Together with the previous result that $x_1^*(\theta'|f) \geq x_1^*(\theta|f)$, we conclude that $x_1^*(\theta'|f) > x_1^*(\theta|f)$. Again, the proof is identical to show that $x_1^*(\theta'|f) < x_1^*(\theta|f)$ if f is strictly decreasing.

□

The above results imply that the single-crossing condition on $\tilde{u}'$ is sufficient, in that it ensures that θ' responds more to the change in incentives than θ. We next show that it is also a necessary condition for a general f function. Formally, we have the following theorem:

Thm B.3. *Suppose $\tilde{u}(x_1,\theta)$ does not have the single crossing condition on $\tilde{u}'$. Then there exists a strictly increasing function such that $x_1^*(\theta'|f) \leq x_1^*(\theta|f)$ or there exists a strictly decreasing function such that $x_1^*(\theta'|f) \geq x_1^*(\theta|f)$.*

Proof. We will initially assume that the failure of the single crossing condition on $\tilde{u}'$ occurs by there being some $\tilde{x}_1 > x_1^*$ such that $\tilde{u}'(\tilde{x}_1,\theta') = \tilde{u}'(\tilde{x}_1,\theta)$ for some $\theta' > \theta$ with $x_1^*(\theta) = x_1^*(\theta') \equiv x_1^*$. We will then show that there exists a strictly increasing differentiable $f(x_1)$ such that $x_1^*(\theta|f) = x_1^*(\theta'|f) = \tilde{x}_1$.

Specifically, for $f(x_1) = \Delta x_1$ with $\Delta = -\tilde{u}'(\tilde{x}_1,\theta)$, we get that $\tilde{u}'(\tilde{x}_1,\theta) + f'(\tilde{x}_1) = \tilde{u}'(\tilde{x}_1,\theta') + f'(\tilde{x}_1) = 0$. From the assumption that $u(x,\theta)$ is concave in x and $\mathcal{E}$ is convex, $\tilde{x}_1$ is the optimal choice of x_1 for both θ' and θ under the added incentive $f(x_1)$ and so $x_1^*(\theta|f) = x_1^*(\theta'|f) = \tilde{x}_1$.

□

C Value-Added Estimation

In this appendix, we describe the different forms of value-added we use in the paper and how we estimate them. Our estimation procedure follows Mulhern and Opper (2021), although Mulhern and Opper (2021) does not control for experience in their estimates.

C.1 Residualizing Outcomes

Let i index students, j index teachers, c index classrooms, and t index years. Let $\tau()$ be a function that describes when an outcome is realized. For contemporaneous outcomes, $\tau(k) = 0$, while for outcomes realized in the future, like next year's test scores, $\tau(k) > 0$. Our statistical model of outcomes, for a specific subject-level, is:

$$y_{i,t+\tau} = \Lambda X'_{it} + \sum_{e'} \rho_{e'} \mathbb{1}\{e_{jt} = e'\} + \mu_{jt} + \nu_{ct} + \phi_{c',t+1}\mathbb{1}(\tau \geq 1) + \phi_{c'',t+2}\mathbb{1}(\tau = 2) + \epsilon_{it} \quad (33)$$

where e_{jt} is a teacher's experience level (with all teachers with six or more years of prior experience grouped into one level).

We have 4 types of outcomes:

1. **Targeted outcome**: test scores in year t
2. **Untargeted outcomes**: test scores in year $t+1$, test scores in year $t+2$, attendance rate in year t, attendance rate in year $t+1$, grades in tested subject in year t, grades in tested subject in year $t+1$, grades in untested subjects in year t, grades in untested subjects in year $t+1$
3. **Index of untargeted outcome**: an index of the above outcomes (constructed below)
4. **Long-term outcome**: whether the student graduates high school on-time

For ease of exposition, label the four outcomes (at the student-level) y^1_{it}, $\vec{y}^2_{it}$, y^3_{it}, y^4_{it}.

In a first step, we standardize each outcome in y^1_{it} and $\vec{y}^2_{it}$ to have mean 0 and standard deviation 1 for each grade-year in NYC.

We then residualize outcomes in y^1_{it}, $\vec{y}^2_{it}$, and y^4_{it} by regressing them on a set of observable characteristics and teacher fixed effects:

$$y_{i,t+\tau} = \Lambda X'_{it} + \sum_{e'} \rho_{e'} \mathbb{1}\{e_{jt} = e'\} + \mu_j + \upsilon_{it}. \quad (34)$$

where $\upsilon_{it} = \mu_{jt} - \mu_j + \nu_{ct} + \phi_{c',t+1}\mathbb{1}(\tau \geq 1) + \phi_{c'',t+2}\mathbb{1}(\tau = 2) + \epsilon_{it}$. We run separate regressions for each outcome, subject (math or ELA), and level (elementary or middle) combination.

We let the set of controls, X_{it}', vary by outcome. For all outcomes, we include year dummy variables and indicators for whether the student receives free or reduced price lunch and whether the student is an English language learner, male, Black, Hispanic, and Asian. For lagged outcomes we use:

- Cubic polynomials in $t-1$ test scores for each subject – used for test scores in t, $t+1$, $t+2$, subject grades in t and $t+1$, untested subject grades in t and $t+1$, graduation
- Cubic polynomial in $t-1$ attendance rate – used for attendance rate in t and $t+1$.

For each student, we construct two residuals:

1. $\hat{\upsilon}_{it}^1 = y_{i,t+\tau} - \hat{\Lambda} X_{it}' - \sum_{e'} \hat{\rho}_{e'} \mathbb{1}\{e_{jt} = e'\}$
2. $\hat{\upsilon}_{it}^2 = y_{i,t+\tau} - \hat{\Lambda} X_{it}'$

The residuals differ in whether the teacher's experience effects are included.

C.2 Constructing Measures of Teacher Output in Each Year

We then construct two (noisy) measures of a teacher's output in year t (for each subject-level) by taking the mean of the two student residuals over each teacher-year combination:

1. $\hat{\mu}_{jkt}^1 = \frac{1}{N_{jkt}} \sum_{i \in \mathcal{I}_{jkt}} \hat{\upsilon}_{ikt}^1$
2. $\hat{\mu}_{jkt}^2 = \frac{1}{N_{jkt}} \sum_{i \in \mathcal{I}_{jkt}} \hat{\upsilon}_{ikt}^2$

where $\mathcal{I}_{jkt}$ is the set of N_{jkt} students with outcome k who are taught by j in year t. We construct these measures for each outcome in y_{it}^1, $\vec{y}_{it}^2$, and y_{it}^4.

For analysis in Section V we use the measure that includes experience effects ($\hat{\mu}_{jkt}^2$) when it is the outcome variable. The exception is Figure 6, where we use the version without experience effects ($\hat{\mu}_{jkt}^1$) to show the flatness of the curve for unexposed cohorts. For analysis in Sections VI and VII, we use the version without experience effects ($\hat{\mu}_{jkt}^1$) as the outcome because we project it onto shrunken value-added measures that exclude the experience profile.

C.3 Constructing Forecasts of Teacher Output

The prior measures are noisy estimates of a teacher's realized output in a given year. For classifying teachers according to their unincentivized output, we construct forecasts that incorporate data from multiple years. We construct forecasts for each outcome in y_{it}^1 and $\vec{y}_{it}^2$.

We follow Mulhern and Opper (2021) and refer there for the details. The key estimation points are:

- The estimates are from a joint Empirical Bayes procedure where the estimates are shrunken jointly.
- We estimate using data from unincentivized periods only. We produce estimates for all years as if they were unincentivized (even if they were actually incentivized).
- We estimate using the non-experience residuals ($\hat{\mu}^1_{jkt}$).
- We allow for the non-experience component of a teacher's effect to drift over time. We let drift rates vary depending on the difference in years between measures, where we estimate a constant drift rate for year differences at least 3.
- We keep teacher-subject-levels with at least ten students with an outcome.
- In forecasting a teacher's output in year t, we use data from all years except year t (i.e., a leave-out estimator).
- For each jt, some of the output measures may be missing. In these cases, we predict the missing measures with forecasts based on the non-missing measures. Specifically, we estimate a separate joint Empirical Bayes procedure for each combination of non-missing measures and use it to forecast the missing measures. The identifying assumption is that the missingness is random conditional on the forecast of the non-missing measures.
- Our inclusion of $\phi_{c',t+1}\mathbb{1}(\tau \geq 1) + \phi_{c'',t+2}\mathbb{1}(\tau = 2)$ in the model means that the correlation structure of a teacher's students' residuals varies with the overlap in their classes in $t+1$ and $t+2$. We incorporate this in constructing the forecasts.

We denote these forecasts as $\tilde{\mu}_{jkt}$ and use them in Sections VI and VII. The exception is the first four columns in Table 8, where we show how treatment effects vary with heterogeneity in *only* the targeted or untargeted forecasts, rather than jointly. For these columns, we construct forecasts from Emprical Bayes procedures that only include the relevant outcomes (y^1_{it} or $\vec{y}^2_{it}$).

C.4 Constructing the Index of Untargeted Output

We create the untargeted index by anchoring the measures to their relative predictiveness of whether a student graduates from high school:

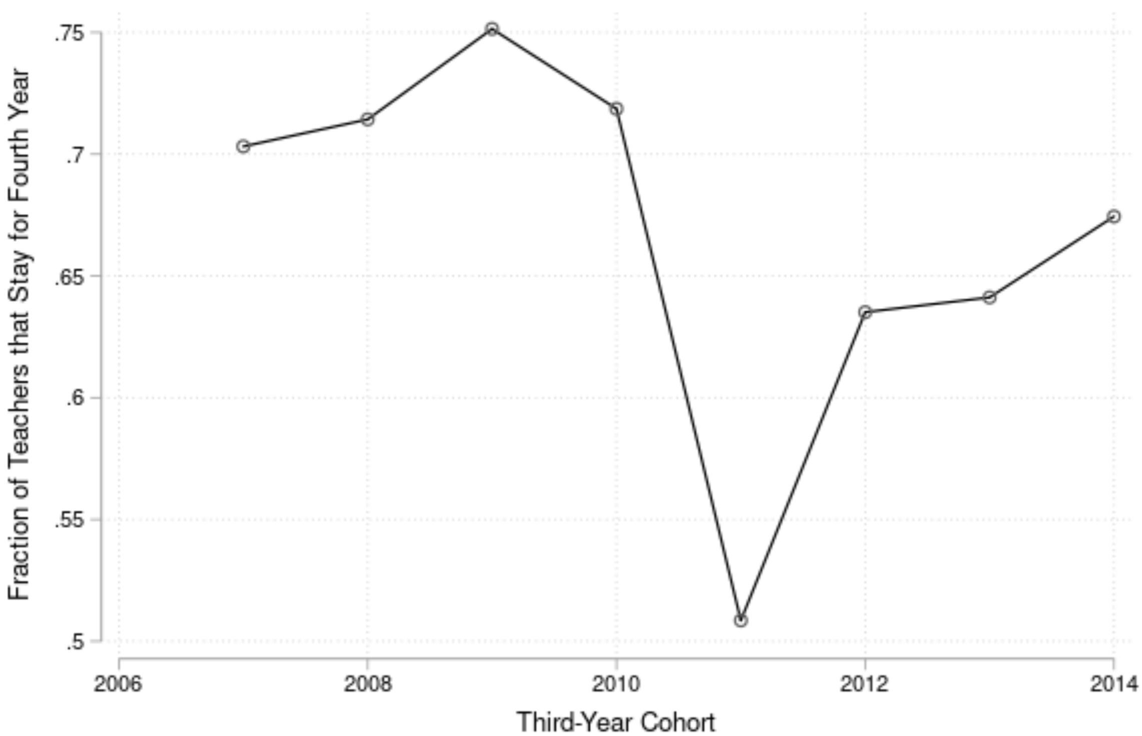
$$\hat{\mu}^1_{jlt} = \omega'\tilde{\mu}_{jt} + \upsilon_{ij}, \tag{35}$$

where l corresponds to the graduation outcome and ω is a vector of anchoring weights. We estimate using data from the unincentivized period.

We use the estimated weights to construct two measures: targeted output ($\tilde{\mu}^T_{jt} = \tilde{\mu}_{j1t}$) and an index of untargeted output ($\tilde{\mu}^U_{jt} = \frac{1}{\hat{\omega}_1}\sum_{k=2}^{K}\hat{\omega}_k\tilde{\mu}_{jkt}$). We also apply these weights to the unshrunken measures for further indices, $\hat{\mu}^T_{jt} = \hat{\mu}_{jkt}$ and $\hat{\mu}^U_{jt} = \frac{1}{\hat{\omega}_1}\sum_{k=2}^{K}\hat{\omega}_k\hat{\mu}_{jkt}$.

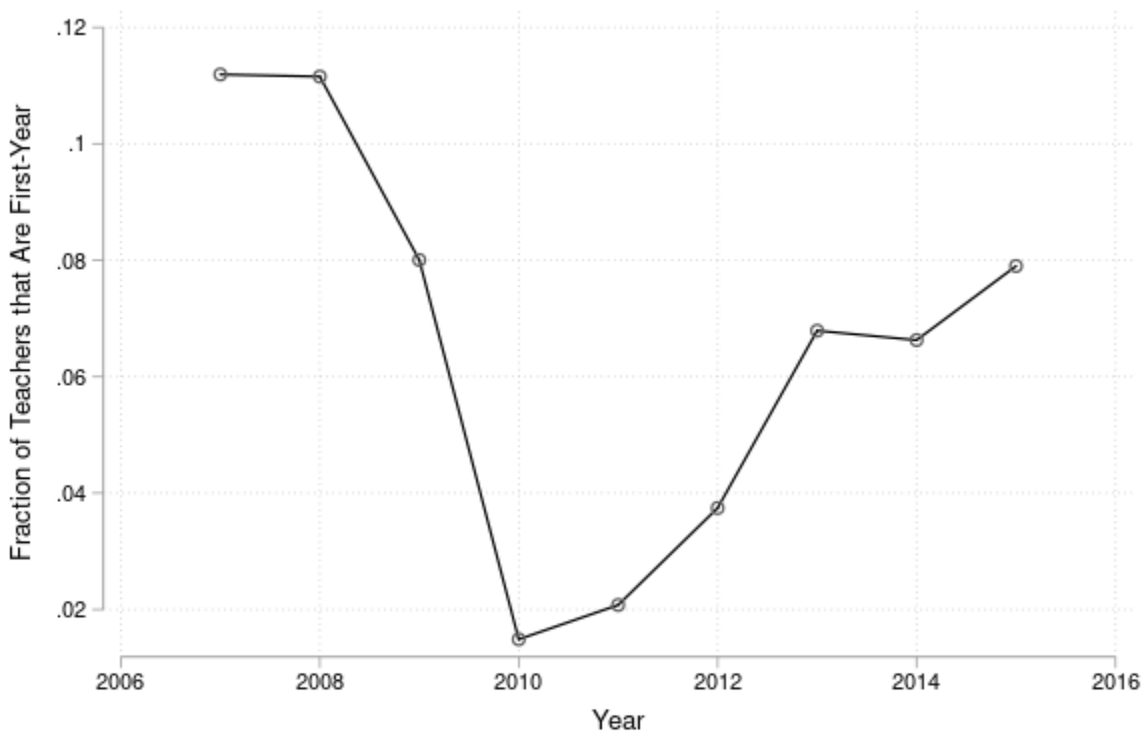
D Appendix Figures

Figure A1: Teachers' Persistence to Fourth Year of Teaching



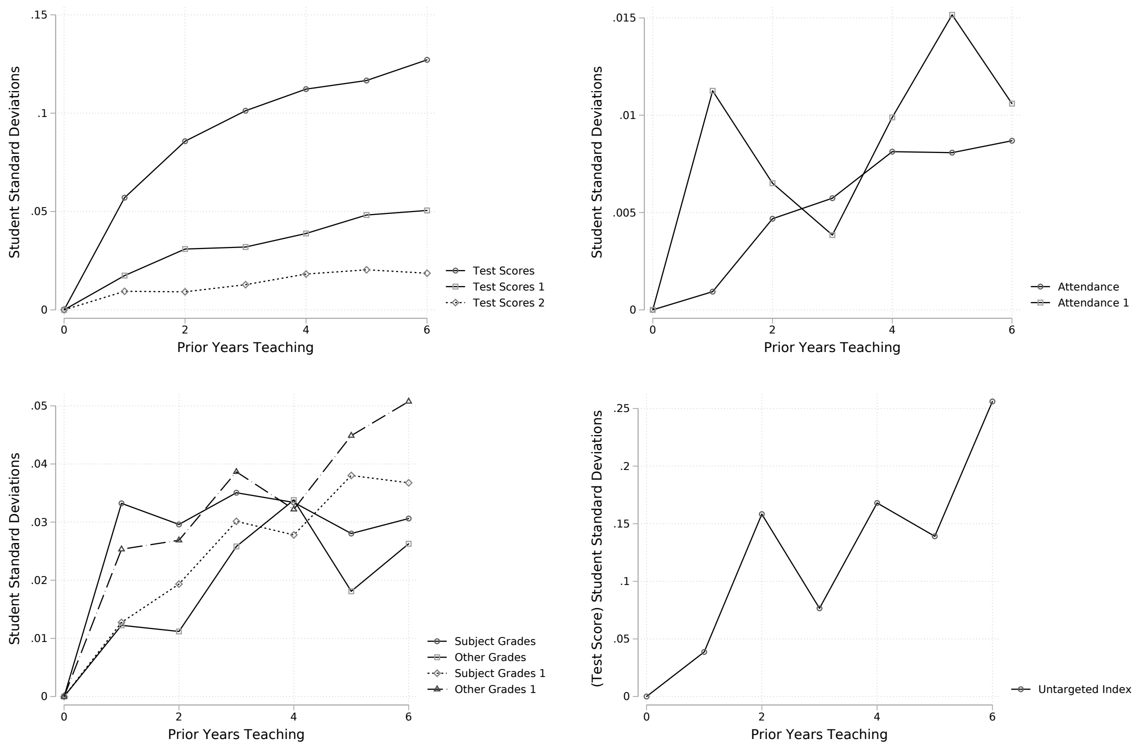
Note: This figure shows the fraction of teachers who were in the district in their third year of teaching who remain in the district their fourth year. The x-axis classifies cohorts based on the academic year when their third year of experience occurred.

Figure A2: New Teachers' Fraction of Workforce



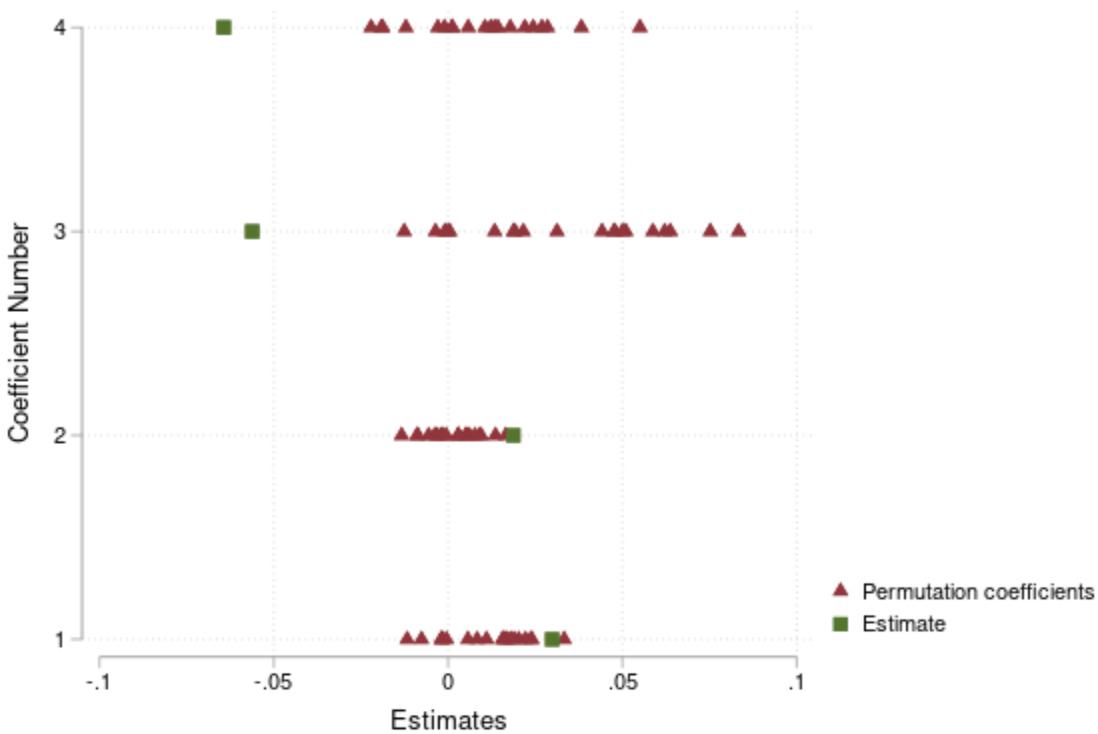
Note: This figure shows the fraction of each year's teacher workforce that first-year teachers comprise in NYC.

Figure A3: Experience profiles



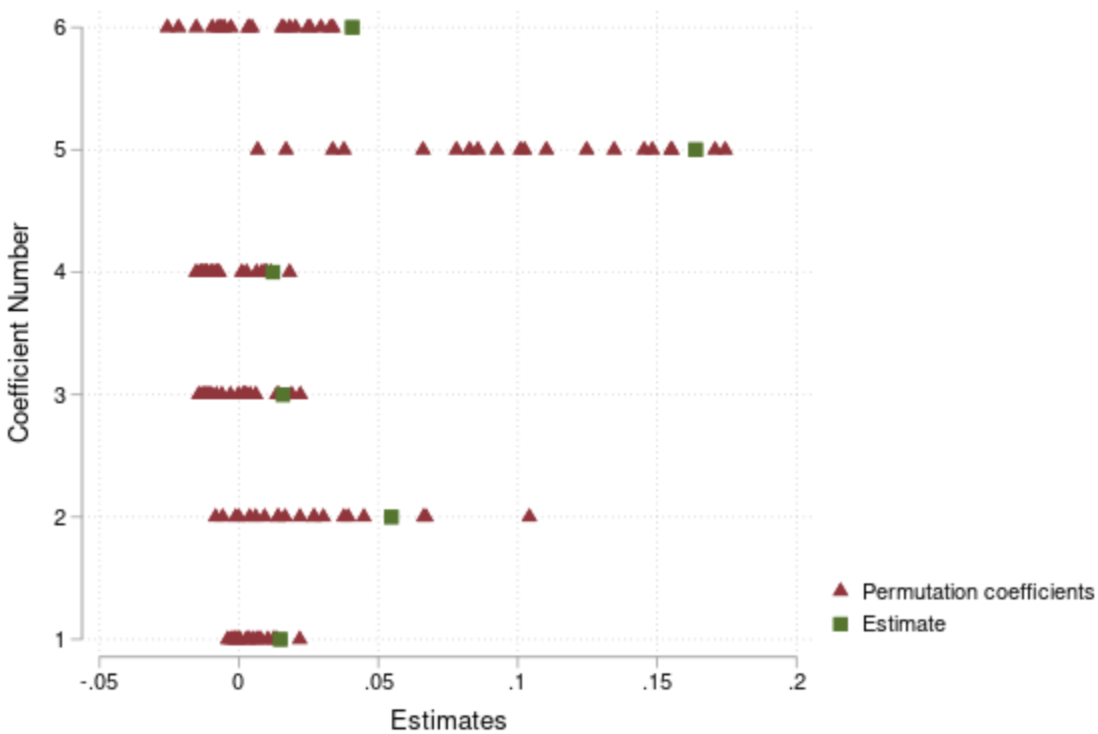
Note: This figure shows the estimated experience profile for our measures, where output in the first year of teaching is normalized to 0 and teachers with six or more years of experience are grouped into one category. The top left panel shows the score measures (current and future), the top right panel shows the attendance measures, the bottom left panel shows the grades measures, and the bottom right panel shows the untargeted index. In the first three panels, units are student standard deviations on each measure. In the bottom right panel, the units are student standard deviations on test scores.

Figure A4: Permutation Tests for Table 4 Estimates



Note: This figure shows permutation tests for the coefficients on "Incentive" in Table 4. The labeled rows correspond to the columns in Table 4. For each test, we assign the policy change to a different set of cohorts. In the correctly specified timing, the 2010 policy affects the cohorts with 0-2 years of prior experience. In the placebo timings, we let the policy affect cohorts with 3-5, 4-6, 5-7, etc. years of prior experience (up to 22-24). We maintain the structure of the policy change and the sample restrictions we impose in our main analysis. The correctly specified regression is labeled with a green square while the placebo estimates are red triangles.

Figure A5: Permutation Tests for Table 8 Estimates



Note: This figure shows permutation tests for the coefficients in columns (5) and (6) of Table 8. Labeled rows (1)-(3) correspond to the estimate in column (5) in Table 8 (from top to bottom). Labeled rows (4)-(6) correspond to the estimate in column (6) in Table 8 (from top to bottom). For each test, we assign the policy change to a different set of cohorts. In the correctly specified timing, the 2010 policy affects the cohorts with 0-2 years of prior experience. In the placebo timings, we let the policy affect cohorts with 3-5, 4-6, 5-7, etc. years of prior experience (up to 22-24). We maintain the structure of the policy change and the sample restrictions we impose in our main analysis. The correctly specified regression is labeled with a green square while the placebo estimates are red triangles.

E Appendix Tables

Table A1: Correlation between Value-Added in t and $t-1$

	Corr b/t VA t and VA t-1	Corr b/t VA t and Test Score VA t-1	Corr b/t VA t and Index VA t-1
Test Score	0.436	0.436	0.161
Untargeted Index	0.560	0.160	0.560
Test Score t+1	0.357	0.196	0.201
Test Score t+2	0.446	0.118	0.249
Attendance	0.553	0.024	-0.121
Attendance t+1	0.265	0.028	0.052
Subject Grades	0.535	0.046	0.121
Other Grades	0.565	0.102	0.344
Subject Grades t+1	0.290	0.045	0.178
Other Grades t+1	0.453	0.074	0.357

This table shows correlations between a teacher's (shrunken) value-added measure in t and various (unshrunken) value-added measures in $t-1$. The columns show correlations with lagged value-added in (1) the same outcome, (2) the targeted measure ("Test Score"), and (3) the index of untargeted measures. We include the targeted measure ("Test Score"), the index of untargeted measures, and each untargeted measure separately.

Table A2: Outcomes' Univariate Relationship to Graduation

	Grad	Grad	Grad	Grad	Grad	Grad	Grad	Grad	Grad
Test Score VA	0.0879***								
	(0.00727)								
Test Score 1 VA		0.0635***							
		(0.00631)							
Test Score 2 VA			0.0716***						
			(0.00798)						
Attendance VA				0.0868***					
				(0.0119)					
Attendance 1 VA					0.0358***				
					(0.00611)				
Subject Grade VA						0.0618***			
						(0.00435)			
Other Grade VA							0.0697***		
							(0.00339)		
Subject Grade 1 VA								0.0923***	
								(0.00617)	
Other Grade 1 VA									0.106***
									(0.00481)
N Teachers	11689	11694	11689	11695	11695	11692	11670	11678	11676
Mean DV	-0.0154	-0.0154	-0.0153	-0.0154	-0.0155	-0.0153	-0.0153	-0.0153	-0.0153
N	32066	32089	32044	32157	32149	32078	31937	31997	31991

This table shows the univariate regression of (the residual of) whether a student graduated from high school on the targeted forecasted measure or on each untargeted forecasted measure. The forecasts come from our multi-year multi-dimensional value-added model that is estimated on unincentivized periods only. Forecasts are constructed for all periods and leave out data from that year. The regression includes only observations from the unincentivized period. In the variable labels, the number indicates the number of years in the future the outcome is realized. All variables are in (separate) standard deviation units.

Table A3: Anchoring Outcomes to Graduation

	Grad
Test Score VA	0.0334***
	(0.0113)
Test Score 1 VA	0.0300*
	(0.0173)
Test Score 2 VA	0.0246*
	(0.0129)
Attendance VA	0.231***
	(0.0211)
Attendance 1 VA	0.0134
	(0.0156)
Subject Grade VA	-0.0149**
	(0.00693)
Other Grade VA	0.0352***
	(0.00549)
Subject Grade 1 VA	-0.0295**
	(0.0142)
Other Grade 1 VA	0.111***
	(0.0108)
N Teachers	11670
Mean DV	-0.0153
N	31937

This table shows the regression of (the residual of) whether a student graduated from high school on the targeted and untargeted forecasted measures. The forecasts come from our multi-year multi-dimensional value-added model that is estimated on unincentivized periods only. Forecasts are constructed for all periods and leave out data from that year. The regression includes only observations from the unincentivized period. In the variable labels, the number indicates the number of years in the future the outcome is realized. All variables are in (separate) standard deviation units.

Table A4: Principal Component Analysis

	First Component	Second Component
Test Score 1 VA	0.156	0.376
Test Score 2 VA	0.151	0.346
Attendance VA	0.015	0.124
Attendance 1 VA	0.051	0.282
Subject Grades VA	0.422	-0.208
Other Grades VA	0.804	-0.350
Subject Grades 1 VA	0.186	0.450
Other Grades 1 VA	0.302	0.525

This table shows the first two components of a Principal Component Analysis on the value-added on each untargeted measure.

Table A5: Effect of Policy Change on Probationary Period Output – By Subject and Level

	Score	Score	Score	Score	Index	Index	Index	Index
Incentive	0.0148	0.0160*	0.0226*	0.00872	-0.0657*	-0.0578	-0.0653*	-0.0669
	(0.0103)	(0.00957)	(0.0130)	(0.00906)	(0.0386)	(0.0387)	(0.0372)	(0.0513)
Fixed Effects	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher
Sample	Math	ELA	Elem	Middle	Math	ELA	Elem	Middle
N Teachers	13184	13868	13389	9909	11961	12494	9036	8030
Mean DV	0.123	0.0810	0.121	0.0542	0.0290	0.0255	-0.0152	0.0915
N	57796	59586	84107	42934	48665	49973	67931	32268

This table shows the causal effect of the tenure policy change on targeted and untargeted output in the probationary period. The columns show the effects in different subsamples, split by the tested subject (Math or ELA) and the level of school (elementary or middle). All regressions include teacher fixed effects. An observation is a teacher-subject-year. Standard errors are clustered by teacher. The sample covers years 2006 to 2014. The teachers with more than 3 years of experience are only included if they finished the standard probationary period before the tenure policy change. All outcome units are test score student standard deviations.

Table A6: Effect of Policy Change on Specific Untargeted Outcomes, Cohort Fixed Effects

	Score 1	Score 2	Attend	Attend 1	Grades	Other Grades	Grades 1	Other Grades 1
Incentive	-0.00572	0.0209	0.00491**	0.000419	0.0139	0.0218	-0.0351*	-0.0188
	(0.00981)	(0.0137)	(0.00212)	(0.00900)	(0.0186)	(0.0159)	(0.0181)	(0.0175)
Teacher FEs	No	No	No	No	No	No	No	No
Mean DV	0.0496	0.0548	0.00417	0.0314	0.0251	-0.00873	0.0398	0.0409
N	110038	87893	132253	110572	36847	48078	60916	74374

This table shows the causal effect of the tenure policy change on individual untargeted (residualized) outcomes in the probationary period. The "1" or "2" in the column headers indicate the measure's number of years into the future. "Grades" are in the tested subject while "Other" grades are in untested subjects. All variables are standardized at the grade-year level to have mean 0 and standard deviation 1 in the full population. All regressions include cohort fixed effects. An observation is a teacher-subject-year. Standard errors are clustered by teacher. The sample covers years 2006 to 2014. The teachers with more than 3 years of experience are only included if they finished the standard probationary period before the tenure policy change.

Table A7: Effect of Policy Change on Specific Untargeted Raw Outcomes, Teacher Fixed Effects

	Score	Score 1	Score 2	Attend	Attend 1	Grades	Other Grades	Grades 1	Other Grades 1
Incentive	0.0135	-0.0252	-0.0151	0.00435	-0.0242*	0.0337	0.000386	-0.0652***	-0.0429*
	(0.0133)	(0.0156)	(0.0192)	(0.0109)	(0.0134)	(0.0214)	(0.0199)	(0.0230)	(0.0219)
Teacher FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	-0.0533	-0.0483	-0.0432	-0.0445	-0.0429	-0.0741	-0.102	-0.0613	-0.0719
N	127678	106647	85244	127480	106756	33217	43512	57508	71371

This table shows the causal effect of the tenure policy change on individual untargeted (non-residualized) outcomes in the probationary period. The "1" or "2" in the column headers indicate the measure's number of years into the future. "Grades" are in the tested subject while "Other" grades are in untested subjects. All variables are standardized at the grade-year level to have mean 0 and standard deviation 1 in the full population. All regressions include teacher fixed effects. An observation is a teacher-subject-year. Standard errors are clustered by teacher. The sample covers years 2006 to 2014. The teachers with more than 3 years of experience are only included if they finished the standard probationary period before the tenure policy change.

Table A8: Effect of Policy Change on Outcomes, Controlling for Next Two Years' Teachers

	Untargeted Index	Untargeted Index	Untargeted Index	Untargeted Index	Score 2	Score 2
Incentive	-0.0749*	-0.0794**	-0.0837**	-0.0837**	-0.00476	-0.00582
	(0.0389)	(0.0389)	(0.0364)	(0.0364)	(0.0135)	(0.0135)
Incentive 1		-0.00913		-0.0102		-0.0267***
		(0.0253)		(0.0198)		(0.00814)
Incentive 2		-0.170***		0.00720		-0.0163**
		(0.0239)		(0.0168)		(0.00740)
Fixed Effects	Cohort	Cohort	Teacher	Teacher	Teacher	Teacher
N Teachers	11324	11324	11324	11324	11202	11202
Mean DV	0.0487	0.0487	0.0487	0.0487	0.0277	0.0277
N	63796	63796	63796	63796	63361	63361

This table shows how our estimates of the effect of the policy change on teachers' untargeted outcomes varies depending on whether we control for the treatment status of the teachers in the two subsequent years. Adjacent columns show the regression with and without these controls, where we restrict the samples to the teachers for whom we can classify the two subsequent teachers. "Score 2" is the test score in year $t + 2$ and is in test score $t + 2$ student standard deviation units while the untargeted index includes all of the untargeted outcomes and is in test sore t student standard deviation units. "Incentive 1" is the fraction of teacher j's students in year t that have an incentivized $t + 1$ teacher, and "Incentive 2" is the fraction that have an incentivized $t + 2$ teacher.

Table A9: Effect of Policy Change on Outcomes, Controlling for Next Year's Teachers

	Score 1	Score 1	Attend 1	Attend 1	Grades 1	Grades 1	Other Grades 1	Other Grades 1
Incentive	-0.0262**	-0.0264**	-0.0348***	-0.0346***	-0.0523**	-0.0508**	-0.0428**	-0.0424**
	(0.0116)	(0.0116)	(0.0114)	(0.0114)	(0.0235)	(0.0235)	(0.0201)	(0.0201)
Incentive 1		-0.0116*		0.00868		0.0522***		0.0169**
		(0.00672)		(0.00625)		(0.0154)		(0.00848)
Fixed Effects	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher
N Teachers	11299	11299	11310	11310	7407	7407	8231	8231
Mean DV	0.0266	0.0266	0.00370	0.00370	0.0234	0.0234	0.0276	0.0276
N	63719	63719	63726	63726	31770	31770	36882	36882

This table shows how our estimates of the effect of the policy change on teachers' untargeted outcomes varies depending on whether we control for the treatment status of the teacher in the subsequent year. Adjacent columns show the regression with and without these controls, where we restrict the samples to the teachers for whom we can classify the subsequent teacher. All outcomes are realized in year $t+1$. Units are student standard deviations for the respective outcome. "Incentive 1" is the fraction of teacher j's students in year t that have an incentivized $t+1$ teacher.

Table A10: Voluntary Attrition by Targeted and Untargeted Value-Added

	Voluntary Attrition
Targeted VA	-0.278***
	(0.0430)
Untargeted VA	-0.0210**
	(0.00914)
N Teachers	5834
Mean DV	0.394
N	8693

This table shows how teachers' voluntary attrition rates vary with their targeted and untargeted forecasted value-added. We consider tenured teachers in their seventh year of teaching in the district and determine whether the teachers left the sample before the end of our data. If so, we label them as voluntary attrition. Targeted and untargeted forecasted value-added are estimated using data from unincentivized periods only and are both in student test score standard deviation units.

Table A11: Output under Different Screening Regimes – 2006 Cohort

	Obs.	Mean Targeted Output	Mean Untargeted Output	Mean Total Output
Teachers' Tenure under Different Responses				
Never Tenured	477	-0.119	-0.771	-0.890
Only Tenured w/o Behavioral Response	50	-0.009	-1.262	-1.271
Only Tenured w/ Behavioral Response	50	-0.041	0.239	0.198
Always Tenured	1,024	0.111	-0.222	-0.111
Tenured Teachers under Different Policies				
Screening w/o Behavioral Response	1,074	0.106	-0.271	-0.165
Screening w/ Behavioral Response	1,074	0.104	-0.201	-0.097
(Infeasible) Screening on Both Dimensions	1,074	0.065	0.010	0.074
Gains Relative to Infeasible First-Best				
Gains (Fraction)				0.285

This table shows the mean output for different groups of teachers and under different policies. The sample is teachers who started in the district in 2006. The top panel splits teachers into four groups based on whether they would receive tenure in a regime without a behavioral response and whether they would receive tenure in a regime with a behavioral response. The middle panel shows the mean output associated with the set of teachers receiving tenure under different policies. The first two policies are screening on the targeted measure, without and with a behavioral response. The last policy is an infeasible policy screening on the sum of output across both dimensions. The final panel shows the fraction of gains the behavioral response achieves relative to the distance between the screening without behavioral response regime and the infeasible policy screening on the sum of output. "Mean Targeted" output is the mean forecasted test score value-added. "Mean Untargeted Output" is the mean forecasted value-added on the untargeted index. "Mean Total Output" is the sum of the mean forecasted value-added across the targeted and untargeted measures. All outcome units are test score student standard deviations.