Multitasking, Learning, and Incentives: A Cautionary Tale

Roland G. Fryer, Jr. and Richard T. Holden*

September 2012

Abstract

We develop a multi-period, multi-task principal-agent model in which the principal knows the mapping from actions to outputs, but the agent does not. The agent can learn about the production function over time by exerting effort and observing output. To test the model's predictions, we conduct a field experiment in fifty Houston public schools, where students, parents, and teachers were rewarded with financial incentives for specific inputs to the education production function. The experimental data is consistent with the model's key predictions, though other explanations are possible. Together, both the theory and experimental evidence serve as a cautionary tale about the efficacy of incentive schemes when agents do not know the production function.

*Special thanks to Terry Grier for his support and leadership during this experiment. We are grateful to Lawrence Katz, Suraj Prasad and Andrei Shleifer for helpful comments and suggestions. Brad Allan, Matt Davis, and Blake Heller provided exceptional research assistance, project management and implementation support. Financial support from the Broad Foundation and the Liemandt Foundation is gratefully acknowledged. Correspondence can be addressed to the authors at: Department of Economics, Harvard University, and NBER [rfryer@fas.harvard.edu] (Fryer); or School of Economics, Australian School of Business, University of New South Wales, and NBER [richard.holden@unsw.edu.au] (Holden).

1 Introduction

Principal-agent models have been used to analyze problems as diverse as executive compensation, regulation, organizational design, entrepreneurship, and accounting.[1] As Kenneth Arrow points out, "economic theory in recent years has recognized that [principal-agent problems] are almost universal in the economy at least as one significant component of almost all transactions" (Arrow, 1986).

In the classic framework, a principal hires an agent to perform a task for her. The agent bears a private cost of taking actions. The principal does not observe the agent's action, rather she observes a noisy measure of it (such as profits). It is this measure that is contractible, and it is assumed that the the agent's cost of effort function, both parties' preferences, and the stochastic mapping from actions to outputs are common knowledge between principal and agent. There have been many important extensions to the basic model (e.g., multitasking and repeated contracting), but it is standard to assume that agents know how their effort affects output.[2]

In many applications, however, it is questionable whether agents know the stochastic mapping from inputs to output. Examples abound in management, finance, and other general contracting problems. In executive management, for instance, it is equivalent to assuming a CEO knows how her actions will impact the collective goals of the board of directors. This requires knowledge of the intensity of differing board-member preferences and how those preferences are aggregated–a complex issue about which board-members themselves are likely better informed through repeated interactions. In education, the standard principal-agent assumptions require that students (or their teachers) know the intricacies of the education production function when econometricians with large data sets and sophisticated statistical techniques are not aware of its functional form.[3]

[1]For classic treatments see Mirrlees 1975, Holmstrom 1979, Grossman and Hart 1983.

[2]See Beaudry (1994), Chade and Silvers (2001), Kaya (2010), and Fryer, Holden, and Lang (2012) for notable exceptions.

[3]Conversely, there are many applications (e.g. computer science, engineering or manufacturing) where the standard assumption seems applicable.

To examine the implications of relaxing this assumption for the design and efficacy of incentive schemes, we begin with a simple 2x2 conceptual apparatus–two periods and two tasks–which is both a simplification and extension of the pioneering work of Holmstrom and Milgrom (1991). In each period, a risk-neutral principal offers a take-it-or-leave-it linear incentive contract to an agent, who, upon accepting the contract, takes two non-verifiable actions which we label "effort." Effort generates a benefit to the principal and is related to an observable (and contractable) performance measure. We assume that an agent's type augments their effort in producing output: higher type agents have higher returns to effort than lower type agents, all else equal.

An important assumption in the model is that the principal and agent have different information about the agent's type before the contracting phase. The principal obtains a private and perfectly revealing signal and hence knows the agents type. The agent observes no private signal, but is assumed to have prior beliefs that are normally distributed. Since the agent does not know her type but the principal does, the contract offered by the principal contains information that is payoff relevant to the agent. We make the stark assumption that the agent *does not* update her prior beliefs after observing the contract.

This is a non-standard assumption for which we provide two justifications. First, it seems internally inconsistent to write down a model in which an agent does not know her own ability but makes sophisticated inferences from the incentive scheme provided. Second, there is a more standard rationale for the agent not updating her beliefs. In Appendix D we show that there exists a pooling equilibrium of the signaling game between principal and agent. In that equilibrium, all types of principal offer the same contract and hence the (fully rational) agent does not update her prior about ability after observing the contract. We leave it to the reader as to whether the behavioral assumption or equilibrium selection argument best rationalizes our assumption that the agent does not update her beliefs when observing the incentive scheme.

The model has four primary predictions. First, incentives for a given task lead to an increase in effort on that task. Second, incentives for a given task lead to a decrease in effort on the non-incentivized task. Further, the decrease in effort on the non-incentivized task can be more or less for higher-type agents relative to lower-type agents, depending on how substitutable those tasks are in the cost of effort function. Our final, and perhaps most distinguishing, theoretical result concerns the persistent effects of changes in incentives due to agents updating about their ability types. We show that when the agent's true ability on a given task is sufficiently low, the learning that comes from the provision of incentives is detrimental to the principal. In the absence of incentives the agent would exert some baseline level of effort due to intrinsic motivation and hence learn "little" about her ability. Providing incentives induces more effort than this and hence more learning about their ability type. When agents discover that they are lower-ability than they previously believed, they exert lower effort in period two for any tasks on which there is a positive incentive slope (as in the case of optimal incentives). The average impact of an incentive contract depends on the distribution of across types, among other things.

To better understand these predictions in a real-world laboratory, we analyze new data from a field experiment conducted in fifty traditionally low-performing public schools in Houston, Texas during the 2010-2011 school year.[4] We provided financial incentives to students, their parents, and their teachers for fifth graders in twenty-five treatment schools.[5] Students received \$2 per math objective mastered in Accelerated Math (AM), a software program that provides practice and assessment of leveled math objectives to complement a primary math curriculum. Students practice AM objectives independently or with assistance on paper worksheets that are scored electronically and verify mastery by taking a computerized test independently at school. Parents also received \$2 for each objective their child mastered and \$20 per parent-teacher conference attended to discuss their student's

[4]The original impetus of the experiment was to study the impact of aligning parent, teacher, and student incentives on student achievement. The two-year evaluation of the experiment led to puzzling findings inconsistent with existing theory. See Fryer (2012).

[5]One may worry that the experiment has incentives for teachers, parents, and students whereas the model has a single agent. If parent and teacher effort has a non-negative effect on student effort, then this is isormorphic to our single agent model with more intense incentives and analogous to the *monitoring intensity principle* in Milgrom and Roberts (1991). Given the lack of impact on direct outcomes in many previous experiments using financial incentives, we chose to align incentives (Angrist and Lavy 2009, Fryer 2011a).

math performance. Teachers earned \$6 for each parent-teacher conference held and up to \$10,100 in performance bonuses for student achievement on standardized tests. In total, we distributed \$51,358 to 46 teachers, \$430,986 to 1,821 parents, and \$393,038 to 1,734 students across the 25 treatment schools.

The experimental results are consistent with the predictions of the model: the good, the bad, and the ugly. Throughout the text we report Intent-to-Treat (ITT) estimates. On outcomes for which we provided direct incentives, there were very large and statistically significant treatment effects. Students in treatment schools mastered 1.087 (0.031) standard deviations (hereafter σ) more math objectives than control students. On average, treatment parents attended almost twice as many parent-teacher conferences as control group parents and were 7.2 percentage points more likely to report checking their child's homework. And, perhaps most important, these behaviors translated into a 0.081σ (0.025) increase in math achievement on Texas's statewide student assessment.

Now, the bad and the ugly: the impact of our incentive scheme on reading achievement (which was not incentivized) is -0.089σ (0.027) – offsetting the positive math effect. And, while higher-achieving students (measured from pre-treatment test scores) seemed to gain from the experiment on nearly every dimension, lower-achieving students had significant and lasting negative treatment effects.

Higher-achieving students master 1.66σ more objectives, have parents who attend two more parent-teacher conferences, have 0.228σ higher standardized math test scores and equal reading scores relative to high-achieving students in non-treated schools. Conversely, lower-achieving students master 0.686σ more objectives, have parents who attend 1.5 more parent-teacher conferences, have equal math test scores and 0.163σ lower reading scores. Put differently, higher-achieving students put in significant effort and were rewarded for that effort in math without a deleterious impact in reading. Lower-achieving students also increased effort on the incentivized task, but did not increase their math scores and their reading scores decreased significantly. These data are compatible with predictions (i) through (iii) of the model.

Consistent with the fourth – and most stark – prediction of the model, higher-achieving students continue to do well, maintaining a positive treatment effect in math and a zero effect in reading, one year after the incentives are taken away. Lower-achieving students, however, exhibit large and statistically significant decreases in both math [-.223σ (0.056)] and reading achievement [-.168σ (0.079)] after the incentives are removed. We argue that this is most likely explained by students learning about their own ability and not decreases in intrinsic motivation. The treatment effect on the latter, gleaned from survey data, is small and statistically insignificant.

Our contribution is three fold. First, we extend the classic multitask principal-agent model to a multi period, multi-type, setting in which the agent does not know the production function, but can learn it over time.[6] Second, we demonstrate, using data from the first experiment designed to align the incentives of parents, teachers, and students on a common performance measure, that the effort substitution problem is larger for low-ability types.[7] Third, we demonstrate persistent negative effects on student test scores on multiple measures – a cautionary tale on the design of incentives when agents do not know the production function.[8]

The next section presents a multi-period, multitasking principal-agent model. Section 3 provides details of the field experiment and its implementation. Section 4 describes the

[6]See Fryer, Holden, and Lang (2012) for a single task model with similar features. Beaudry (1994) also studies a setting where the principal knows the mapping from action to output but the agent does not. In his model there are two types of agent and two possible output levels. Focusing on separating perfect Bayesian equilibria he shows that high types receive a higher base wage and a lower bonus than low types. See also Chade and Silvers (2001) and Kaya (2010). Our also relates to the so-called *informed principal problem* in mechanisms design first analyzed by Myerson (1983) and Maskin and Tirole (1990, 1992). This large literature studies studies the equilibrium choice of mechanisms by a mechanism designer who possess private information. The key difference is that our focus is on a specific environment with hidden actions *after* the contracting stage, rather than on characterizing the set of equilibria in very general hidden information settings. One way to see this difference is that in Maskin and Tirole (1992) actions are observable and verifiable.

[7]There is a growing literature on the use of financial incentives to increase student achievement in primary (Bettinger 2010, Fryer 2011a), secondary (Angrist and Lavy 2009, Fryer 2011a, Kremer, Miguel, and Thornton 2009), and postsecondary (Angrist, Lang, and Oreopoulos 2009, Oosterbeek et al. 2010) education.

[8]Psychologists often warn of the potential negative effects of incentives due to intrinsic motivation. Our model and data suggests a different mechanism: rational, but potentially incorrect, learning about one's type.

data collected, research design, and econometric framework used in the analysis. Section 5 presents estimates of the impact of the treatment on various test score and non-test score outcomes. The final section concludes with a more speculative discussion of the implications of the model and experimental data for the design of incentive schemes. There are three online appendices. Online A provides technical proofs of the propositions detailed in Section 2, along with other mathematical details. Online Appendix B is an implementation supplement that provides details on the timing of our experimental roll-out and critical milestones reached. Online Appendix C is a data appendix that provides details on how we construct our covariates and our samples from the school district administrative files used in our analysis.

2 A Multi-period, Multitasking Model with Learning

2.1 Statement of the problem

In each of two periods, a risk-neutral principal offers a take-it-or-leave-it incentive contract to an agent, who, upon accepting the contract, takes two non-verifiable actions e_1 and e_2. We will typically refer to these actions as *effort*. Each action takes values in $\mathbb{R}_+$, and generates a benefit on task i of $\alpha_i e_i$ to the principal and a performance measure $m_i = \alpha_i e_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma_i^2)$ and is independent of everything else. We will sometimes refer to the level of α_i as the "type" of the agent on task i.

We assume that only the m_i's are contractable, and the principal offers a linear incentive contract of the form $s + b_1 m_1 + b_2 m_2$ that the agent can accept or reject. If the agent accepts she then makes her effort choice(s), the performance measure is realized, and the principal pays the agent according to the contract.

A key assumption of our model is that principal knows the true value of α_1 and α_2, but the agent does not. Before the contract is offered the agent has a prior probability probability distribution $\alpha_i \sim N(\overline{\alpha}_i, \mu_i^2)$. We assume that it is common knowledge between the principal and agent that α does not change over time, and the ϵ_is are independent of each other and i.i.d. over time.

Since the agent does not know the α_is but the principal does, the contract offered by the principal contains information that is payoff relevant to the agent. We make the stark assumption that the agent *does not* update her prior beliefs after observing the contract. This is a non-standard assumption and thus warrants some explanation. As we mentioned in the introduction there is a certain internal inconsistency to an agent who does not know her own ability but makes sophisticated inferences from the incentive scheme provided. But there is also a more standard rationale for the agent not updating. In the appendix, adapting the analysis in Fryer-Holden and Lang (2012) to this setting, we show that there exists a pooling equilibrium of the signaling game between principal and agent. In that equilibrium all types of principal offer the same contract and hence the (fully rational) agent does not update her prior about ability after observing the contract. We leave it to the reader as to whether the behavioral assumption or equilibrium selection argument best rationalizes our assumption that the agent does not update.

We further assume that the agent has preferences that can be represented by a utility function that exhibits constant absolute risk aversion (CARA):

$$u(x,e) = -\exp\left[-\eta\left(x - \frac{1}{2}(e_1^2 + e_2^2) - \delta e_1 e_2\right)\right],$$

where x is the monetary payment she receives. Let $\overline{U}$ be the certainty equivalent of the agent's outside option and normalize this to zero. Notice that the parameter δ (which we assume to be strictly positive) measures the degree of substitutability between the tasks.

Finally, we assume that the agent is myopic and unable to borrow, and we normalized the common discount factor to 1.

2.2 Solving the model

2.2.1 One period, symmetric information

We now solve the benchmark case where the agent also knows α. Given the exponential utility function and normal noise standard calculation imply that the agent receives certainty equivalent

$$CE = \sum_{i=1}^{2} b_i \alpha_i e_i + s - \frac{1}{2}(e_1^2 + e_2^2) - \delta e_1 e_2 - \sum_{i=1}^{2} \frac{\eta}{2} b_i^2 \sigma_i^2.$$

Therefore, the principal's problem becomes

$$\max_{b_1,b_2,s,e_1,e_2} \left\{ \sum_{i=1}^{2} (\alpha_i - b_i) e_i - s \right\}$$

$$\text{subject to} \quad e_1, e_2 \in \text{argmax}_{\tilde{e}_1,\tilde{e}_2} \left\{ \sum_{i=1}^{2} b_i \alpha_i \tilde{e}_i + s - \frac{1}{2}\left(\tilde{e}_1^2 + \tilde{e}_2^2\right) - \delta \tilde{e}_1 \tilde{e}_2 - \sum_{i=1}^{2} \frac{\eta}{2} b_i^2 \sigma_i^2 \right\}$$

$$\sum_{i=1}^{2} b_i \alpha_i e_i + s - \frac{1}{2}(e_1^2 + e_2^2) - \delta e_1 e_2 - \sum_{i=1}^{2} \frac{\eta}{2} b_i^2 \sigma_i^2 \geq \overline{U}.$$

The first constraint is the agent's incentive compatibility (IC) constraint, ensuring that the efforts that the principal designs the incentive scheme to elicit are in fact optimal for the agent. The second is the agent's individual rationality (IR) constraint, ensuring that the agent receives at least her outside option in expectation and hence is willing to accept the contract offered by the principal.

The solution to the agent's optimization problem[9] is

$$e_1 = \alpha_1 b_1 - \delta e_2,$$

$$e_2 = \alpha_2 b_2 - \delta e_1.$$

[9]Note that the first-order approach is valid in this setting–that is the second-order conditions for the agent's problem are satisfied.

Solving simultaneously yields

$$e_1^* = \frac{b_1\alpha_1 - \delta b_2\alpha_2}{1-\delta^2}, \tag{1}$$

$$e_2^* = \frac{b_2\alpha_2 - \delta b_1\alpha_1}{1-\delta^2}. \tag{2}$$

Substituting these and the s that makes the agent's participation constraint binding into the principal's objective function and taking first-order conditions for b_1 and b_2 yields the following unconstrained problem.[10]

$$\begin{aligned}\max_{b_1,b_2}\Bigg\{&\left(\frac{b_1\alpha_1(1-\delta)+b_2\alpha_2(1-\delta)}{1-\delta^2}\right)-\delta\left(\frac{b_1\alpha_1-\delta b_2\alpha_2}{1-\delta^2}\right)\left(\frac{b_2\alpha_2-\delta b_1\alpha_1}{1-\delta^2}\right)\\ -\frac{1}{2}&\left(\left(\frac{b_1\alpha_1-\delta b_2\alpha_2}{1-\delta^2}\right)^2+\left(\frac{b_2\alpha_2-\delta b_1\alpha_1}{1-\delta^2}\right)^2\right)-\frac{\eta}{2}\left(b_1^2\sigma_1^2+b_2^2\sigma_2^2\right).\Bigg\}\end{aligned} \tag{3}$$

Taking first-order conditions and solving simultaneously yields the equilibrium incentive slopes (b_1^*, b_2^*),

$$b_1^* = \frac{(\alpha_1)^2+(1-\delta)\eta\sigma_2^2}{(\alpha_1)^2+\eta(\sigma_2^2+\sigma_1^2)+\eta^2\sigma_1^2\sigma_2^2(1-\delta^2)}, \tag{4}$$

$$b_2^* = \frac{(\alpha_2)^2+(1-\delta)\eta\sigma_1^2}{(\alpha_2)^2+\eta(\sigma_1^2+\sigma_2^2)+\eta^2\sigma_2^2\sigma_1^2(1-\delta^2)}. \tag{5}$$

2.2.2 One period, asymmetric information

The analysis proceeds as above, but now the certainty equivalent now must account for the risk imposed on the agent because of the uncertainty about α, and that the agent must take an expectation over α when assessing her expected payment.

$$CE = \sum_{i=1}^{2} b_i\overline{\alpha}_i e_i + s - \frac{1}{2}(e_1^2+e_2^2) - \delta e_1 e_2 - \sum_{i=1}^{2}\frac{\eta}{2}b_i^2\left(\sigma_i^2+e^2\mu_i^2\right).$$

The principal's problem has the same objective function, but the IR and IC constraint

[10] The participation constraint must at the optimum, otherwise the fixed payment s could be reduced and improve the principal's payoff with affective the agent's incentives of payoff.

now take account of the uncertainty about α.

$$\begin{aligned}
&\max_{b_1,b_2,s,e_1,e_2} \left\{ \sum_{i=1}^{2} (\alpha_i - b_i) e_i - s \right\} \\
\text{subject to} \quad & e_1, e_2 \in \operatorname{argmax}_{\tilde{e}_1,\tilde{e}_2} \left\{ \sum_{i=1}^{2} b_i \overline{\alpha}_i \tilde{e}_i + s - \frac{1}{2}\left(\tilde{e}_1^2 + \tilde{e}_2^2\right) - \delta \tilde{e}_1 \tilde{e}_2 - \sum_{i=1}^{2} \frac{\eta}{2} b_i^2 \left(\sigma_i^2 + e^2 \mu_i^2\right) \right\} \\
& \sum_{i=1}^{2} b_i \overline{\alpha}_i e_i + s - \frac{1}{2}(e_1^2 + e_2^2) - \delta e_1 e_2 - \sum_{i=1}^{2} \frac{\eta}{2} b_i^2 \left(\sigma_i^2 + e^2 \mu_i^2\right) \geq \overline{U}.
\end{aligned}$$

The solution to the agent's optimization problem is now

$$\begin{aligned}
e_1 &= \overline{\alpha}_1 b_1 - \frac{1}{2} b_1^2 \eta \mu_1^2 - \delta e_2 - e_1, \\
e_2 &= \overline{\alpha}_2 b_2 - \frac{1}{2} b_2^2 \eta \mu_2^2 - \delta e_1 - e_2.
\end{aligned}$$

Solving simultaneously we have

$$e_1^* = \frac{2\overline{\alpha}_1 b_1 - 2\overline{\alpha}_2 b_2 \delta + b_2^2 \delta \eta \mu_2^2 - b_1^2 \eta \mu_1^2}{2\left(1 - \delta^2\right)}, \tag{6}$$

$$e_2^* = \frac{2\overline{\alpha}_2 b_2 - 2\overline{\alpha}_1 b_1 \delta + b_1^2 \delta \eta \mu_1^2 - b_2^2 \eta \mu_2^2}{2\left(1 - \delta^2\right).} \tag{7}$$

Proceeding as before, the principal's unconstrained problem is

$$
\begin{aligned}
&\max_{b_1,b_2}\Bigg\{-2\Bigg(\frac{1}{2}b_1^2\eta\left(\sigma_1^2-\frac{\mu_1^2\left(-2\overline{\alpha}_2b_2\delta+2\overline{\alpha}_1b_1+b_2^2\delta\eta\mu_2^2+b_1^2(-\eta)\mu_1^2\right)}{2\left(\delta^2-1\right)}\right)\\
&+\frac{1}{2}b_2^2\eta\left(\sigma_2^2-\frac{\mu_2^2\left(-2\overline{\alpha}_1b_1\delta+2\overline{\alpha}_2b_2+b_1^2\delta\eta\mu_1^2-b_2^2\eta\mu_2^2\right)}{2\left(\delta^2-1\right)}\right)\\
&+\frac{\overline{\alpha}_1b_1\left(-2\overline{\alpha}_2b_2\delta+2\overline{\alpha}_1b_1+b_2^2\delta\eta\mu_2^2+b_1^2(-\eta)\mu_1^2\right)}{2\left(\delta^2-1\right)}+\frac{\overline{\alpha}_2b_2\left(-2\overline{\alpha}_1b_1\delta+2\overline{\alpha}_2b_2+b_1^2\delta\eta\mu_1^2-b_2^2\eta\mu_2^2\right)}{2\left(\delta^2-1\right)}\\
&+\frac{\delta\left(-2\overline{\alpha}_1b_1\delta+2\overline{\alpha}_2b_2+b_1^2\delta\eta\mu_1^2-b_2^2\eta\mu_2^2\right)\left(-2\overline{\alpha}_2b_2\delta+2\overline{\alpha}_1b_1+b_2^2\delta\eta\mu_2^2+b_1^2(-\eta)\mu_1^2\right)}{4\left(\delta^2-1\right)^2}\\
&+\frac{1}{2}\left(\frac{\left(-2\overline{\alpha}_1b_1\delta+2\overline{\alpha}_2b_2+b_1^2\delta\eta\mu_1^2-b_2^2\eta\mu_2^2\right)^2}{4\left(\delta^2-1\right)^2}+\frac{\left(-2\overline{\alpha}_2b_2\delta+2\overline{\alpha}_1b_1+b_2^2\delta\eta\mu_2^2+b_1^2(-\eta)\mu_1^2\right)^2}{4\left(\delta^2-1\right)^2}\right)\Bigg)\\
&-\frac{\left(\overline{\alpha}_2-b_2\right)\left(-2\overline{\alpha}_1b_1\delta+2\overline{\alpha}_2b_2+b_1^2\delta\eta\mu_1^2-b_2^2\eta\mu_2^2\right)}{2\left(\delta^2-1\right)}\\
&-\frac{\left(\overline{\alpha}_1-b_1\right)\left(-2\overline{\alpha}_2b_2\delta+2\overline{\alpha}_1b_1+b_2^2\delta\eta\mu_2^2+b_1^2(-\eta)\mu_1^2\right)}{2\left(\delta^2-1\right)}\Bigg\}
\end{aligned}
\tag{8}
$$

It is unimportant–although possible–to obtain closed-form solutions for the incentive slopes as we can proceed with our analysis using monotone comparative statics on the unconstrained objective function.[11]

It is worth noting, however, that it is the agent's beliefs about the αs, not the true values, that end up entering into the slope of the incentive scheme. The principal must satisfy the IC and IR constraints of the agent, and these depend on the agent's beliefs. This is precisely why manipulating the agent's beliefs can affect effort.

2.2.3 Agent updating

Now consider the two-period problem that the principal faces. She cannot change the agent's actions in period 1, but after period one the agent updates her belief about α_1 and α_2 based on the outputs her actions generated. Thus, the choice of b_1 and b_2 in period 1 can affect the agent's actions in period two through these beliefs. After taking actions (e_1^1, e_2^1) (superscripts index the period) and observing outputs (m_1^1, m_2^1) the agent's posterior belief about her ability on task i are:

[11] By using monotone methods we do not need to address the issue of whether the principal's problem is now convex when there is asymmetric information about α.

$$E[\alpha_i|m_i^1] = \bar{\alpha}_i \left(\frac{\sigma_i^2}{\mu_i^2+\sigma_i^2}\right) + \frac{2(\alpha_i-\overline{\alpha}_i)b_i - 2(\alpha_{-i}-\overline{\alpha}_{-i})b_{-i}\delta + b_{-i}^2\delta\eta\mu_2^2 - b_i^2\eta\mu_1^2}{2\left(1-\delta^2\right)}\left(\frac{\mu_i^2}{\mu_i^2+\sigma_i^2}\right). \tag{9}$$

In forming her posterior the agent puts some weight on her prior, and some weight on first period output, which depends on her effort and her true ability. This obviously bears strong similarities to the classic career concerns model of Holmstrom (1982) in terms of the way the agent updates about her ability (see also, very closely related, Dewatripont-Jewitt-Tirole (1999)).

Other than the well-know role that the signal-to-noise ratio plays, notice that the agent's posterior is increasing in the difference between α_i and her expectation of it $\overline{\alpha}_i$. This will play a key role, since the principal can increase expected output by using more intense incentives in period 1. Thus she can to some degree control of surprised the agent is, although this comes at a cost, because the IR constraint must be satisfied and that depends on the agent's subjective first-period belief of her ability.

2.3 Results

To ground ideas, the first and most basic result is a version of the *effort substitution problem* of Holmstrom and Milgrom (1991). When it becomes optimal for incentives to become larger on task i, due to a change in primitives in period 1, it is necessarily optimal for incentives to become smaller on task $j \neq i$.

Proposition 1 *Consider the one period model. Fix σ_i^2, σ_j^2 and associated $b_i^*(\sigma_i^2, \sigma_i^2)$ and consider $\hat{\sigma}_j^2 < \sigma_j^2$. Then $b_i^*(\sigma_i, \hat{\sigma}_j^2) < b_i^*(\sigma_i, \sigma_j^2)$ and $b_j^*(\sigma_i, \hat{\sigma}_j^2) > b_i^*(\sigma_i, \sigma_j^2)$.*

Our next set of results speak to interventions in incentive schemes. Here we have in mind a situation where incentives on one task are increased in period one, relative to some benchmark level. A natural interpretation is that the benchmark level is the optimal level and the increase comes from some change in (exogenous) primitives in our model, such as the benefit to the principal or the precision of measurement of output. The following result says that an increase in incentives on a given task in period 1: (a) increases effort on that task in period 1 for all types of agent, (b) decreases effort on the other task in period 1 for all types of agent, and (c) may decreases effort on the other task in period 1 *more or less* for higher type agents.

Proposition 2 *Consider the one period model and let the optimal incentive intensities in period 1 be $b^*_{1,1}, b^*_{2,1}$. Now consider a shock to the environment and new optimal incentive intensities $\hat{b}_{1,1}, \hat{b}_{2,1}$ is such that $\hat{b}_{1,1} > b^*_{1,1}$ and $\hat{b}_{2,1} = b^*_{2,1}$. Then (a) $\hat{e}_{1,1} > e^*_{1,1}$, (b) $\hat{e}_{2,1} < e^*_{2,1}$, (c) the sign of $\frac{\partial^2 e_2}{\partial b_1 \partial \bar{\alpha}_1}$ is ambiguous.*

If the "type" of the agents differs not only according to their ability but also to their opportunity cost of time, as measured by δ, the degree of substitutability between tasks, then we can establish a starker result in terms of part (c) of the previous proposition.

Proposition 3 *Consider the one period model and let the optimal incentive intensities in period 1 be $b^*_{1,1}, b^*_{2,1}$. Now consider a shock to the environment and new optimal incentive intensities $\hat{b}_{1,1}, \hat{b}_{2,1}$ is such that $\hat{b}_{1,1} > b^*_{1,1}$ and $\hat{b}_{2,1} = b^*_{2,1}$ and further that the degree of substitutability between tasks changes from δ to δ'. Then there exists a δ' such that $\hat{e}_{1,1} > e^*_{1,1}$, $\hat{e}_{2,1} = e^*_{2,1}$, and that for all $\delta > \delta'$ we have $\hat{e}_{2,1} < e^*_{2,1}$.*

In words, if the concept of ability is enriched to include a measure of opportunity cost of effort in addition to productivity on tasks, then an increase in incentives on one task leads not only to the standard effort substitution problem, but also a differential one across ability types. In fact, it can be the case that some types exhibit no deleterious effect from the effort substitution problem on the second task, while other, lower-ability types do.

A high value of δ implies that raising effort on one task increases the marginal cost of effort on the other task. This is consistent with a setting where there is a third numeraire task (such as leisure) and there is heterogeneity among agents about how they value the numeraire.

Our final results concern the persistent effects of changes in incentives due to agents updating about their types.

Proposition 4 *Let $e_1^*(\alpha_i)$ be agent is effort on task 1 in the absence of incentives (i.e. when b_1=0). Suppose $b_{1,1} > 0$ and $b_{2,1} = 0$ and further that $b_{1,2} = b_{1,2} = 0$. Then there exists $\hat{\alpha_1^*}$ such that for all $\alpha_1^* < \hat{\alpha_1^*} : e_{1,2} < e_1^*$.*

When the agent's true ability on task 1 is sufficiently low the learning that comes from the provision of incentives is detrimental to the principal. In the absence of incentives the agent would exert some baseline level of effort due to intrinsic motivation (in our model literally zero) and hence learn "little" (again, literally zero in our model) about her ability. Providing incentives induces more effort than this and hence more learning about ability. When agents discover that they are lower-ability than they thought they exert lower effort in period two for any tasks on which there is a positive incentive slope (as in the case of optimal incentives).

The fact that there is a cutoff type, above which increased period-one incentives lead to an positive update and below which leads to a negative update stems from the fact that more intense incentives in period 1 lead to a Blackwell-more-informative experiment about agent ability. But Bayes' Rule implies that the expectation of the conditional expectation of ability given period 1 output must equal the unconditional expectation. Thus, when the experiment leads to some agents updating positively about their ability, it must also lead (from an ex ante perspective) to some agents updating negatively.

Suppose now that there is correlation between agents' abilities on the two tasks in the following sense. Suppose α_i^1 and α_i^2 are drawn from distributions F_1 and F_2 respectively and further suppose that these draws are not statistically independent. We will refer to this as abilities being correlated. We then have

Corollary 5 *If abilities are correlated then there exists $\overline{\alpha}$ such that for all $\alpha < \overline{\alpha} : e_{1,2} < e_1^*$ and $e_{2,2} < e_2^*$.*

That is, if agents have correlated abilities incentive that induce learning that have a deleterious effect on one task spill over to the other task. In its most general form, if ρ is the correlation coefficient between α_1^* and α_2^* then larger values of ρ lead to lower effort levels on the second (spillover) task.

Finally, we discuss the optimal period 1 incentive scheme. We can decompose the problem into the costs an benefits of providing additional incentives subject to the IR and IC constraints in period 1.

Divide the agents into to categories: "high types" and "low types". For agent's whose true ability is above the $\hat{\alpha_1^*}$ of Proposition 4 ("high types"), the benefit of increased incentives in period 1 is the additional effort that will be exerted in period 2 as a consequence of the updated beliefs about ability.

The agent's optimal period 2 action on task 1, given her updated beliefs is

$$e_1^2 = \frac{b_1^2 - E[\alpha_1|m_1^1] - \delta b_2^2 E[\alpha_2|m_2^1]}{1-\delta^2},$$

and symmetrically for task 2.

Let us denote the expected increase in profit from this additional effort as $\Pi(e_1^2(e_1^1))$. Note that this includes not only the benefit of additional effort in period 2, but the cost of satisfying the period 2 IR constraint. This is is a net increase in profit. The total expected benefit across all high type agents is then $\int_{\hat{\alpha_1^*}}^{\infty} \Pi(e_1^2(e_1^1))$. For "low types" this "benefit" is negative and given by $\int_{-\infty}^{\hat{\alpha_1^*}} \Pi(e_1^2(e_1^1))$.

The cost of providing additional incentives in period 1 is the same for both classes of types, since they have the same prior belief in period 1. This cost is given by the additional fixed payment s_1^1 required to satisfy the agent's IR constraint. Denote this cost $C(b_1^1, b_2^1)$.[12]

The objective for the principal is thus

$$\max_{b_1^1} \left\{ \int_{\hat{\alpha_1^*}}^{\infty} \Pi(e_1^2(e_1^1)) + \int_{-\infty}^{\hat{\alpha_1^*}} \Pi(e_1^2(e_1^1)) - C(b_1^1, b_2^1) + \sum_{i=1}^{2} (\alpha_i^1 - b_i^1) e_i^1 - s^1 \right\}, \tag{10}$$

[12]The incentive slope on the second task enters because of the quadratic cost function.

and analogously for b_1, both subject to the first period IC and IR constraints. The final term is the expected profit in period 1. As before, solving the first period IR constraint for s and the replacing the first period IC constraint with the agent's FOC we have an unconstrained problem for the principal. Taking first-order conditions and solving simultaneously leads to the optimal period 1 incentive slopes.

This is, of course, rather abstract, and we do not pursue closed-form solutions here as they are not relevant for our purpose. The salient point to note, however, is that the distribution of types is a crucial determinant of the optimal incentive scheme. This is because, when agents do not know their types and the principal is constrained to offer a single contract to all types there is a tradeoff between the benefits from "teaching" higher types their true ability, and a cost of teaching lower types theirs. We return to this in our concluding remarks.

3 Program Details

Houston Independent School District (HISD) is the seventh largest school district in the nation with 202,773 students. Eighty-eight percent of HISD students are black or Hispanic. Roughly 80 percent of all students are eligible for free or reduced-price lunch and roughly 30 percent of students have limited English proficiency.

Table 1 provides a bird's-eye view of the demonstration project. To begin the field experiment, we followed standard protocols. First, we garnered support from the district superintendent and other key district personnel. Following their approval, a letter was sent to seventy-one elementary school principals who had the lowest math performance in the school district in the previous year. In August 2010, we met with interested principals to discuss the details of the experiment and provided a five day window for schools to opt into the randomization. Schools that signed up to participate serve as the basis for our matched-pair randomization. All randomization was done at the school level. Prior to the randomization, all teachers in the experimental group signed a non-binding commitment form vowing to use the Accelerated Math curriculum to supplement and complement their regular math instruction and indicating their intention to give all students a chance to master Accelerated Math objectives on a regular basis regardless of their treatment assignment.[13] After treatment and control schools were chosen, treatment schools were alerted that they would participate in the incentive program. Control schools were informed that they were not chosen, but they would still receive the Accelerated Math software – just not the financial incentives.[14] HISD decided that students and parents at selected schools would be automatically enrolled in the program. Parents could choose not to participate and return a signed opt-out form at any point during the school year.[15] HISD also decided that students and parents were required to participate jointly: students could not participate without their parents and vice versa. Students and parents received their first incentive payments on October 20, 2010 and their last incentive payment on June 1, 2011; teachers received incentives with their regular paychecks.[16]

Table 2 describes differences between schools that signed up to participate and other elementary schools in HISD with at least one fifth grade class across a set of covariates. Experimental schools have a higher concentration of minority students and teachers with low-value added on math scores. All other covariates are statistically similar.

A. Students

Students begin the program year by taking an initial diagnostic assessment to measure mastery of math concepts, after which AM creates customized practice assignments that focus specifically on areas of weakness. Teachers assign these customized practice sheets, and students are then able to print the assignments and take them home to work on (with or without their parents). Each assignment has six questions, and students must answer at least

[13]This was the strongest compliance mechanism that the Harvard Institutional Review Board would allow for this experiment. Teachers whose data revealed that they were not using the program were targeted with reminders to use the curriculum to supplement and complement their normal classroom instruction. All such directives were non-binding and did not affect district performance assessments or bonuses.

[14]Schools varied in how they provided computer access to students (e.g. some schools had laptop carts, others had desktops in each classroom, and others had shared computer labs), but there was no known systematic variation between treatment and control.

[15]Less than 1%, 2 out of 1695 parents opted out of the program.

[16]In the few cases in which parents were school district employees, we paid them separately from their paycheck.

five questions correctly to receive credit.[17] After students scan their completed assignments into AM, the assignments are graded electronically. Teachers then administer an AM test that serves as the basis for potential rewards; students are given credit for official mastery by answering at least four out of five questions correctly. Students earned $2 for every objective mastered in this way. Students who mastered 200 objectives were declared "Math Stars" and received a $100 completion bonus with a special certificate.[18]

B. Parents

Parents of children at treatment schools earned up to $160 for attending eight parent-teacher review sessions ($20/each) in which teachers presented student progress using Accelerated Math Progress Monitoring dashboards. Appendix Figure 1 provides a typical example. Parents and teachers were both required to sign and submit the student progress dashboards and submit them to their school's Math Stars coordinator in order to receive credit. Additionally, parents earned $2 for their child's mastery of each AM curriculum objective, so long as they attended at least one conference with their child's teacher. This requirement also applied retroactively: if a parent first attended a conference during the final pay period, the parent would receive a lump sum of $2 for each objective mastered by their child to date. Parents were not instructed on how to help their children complete math worksheets.

[17] Accelerated Math does not have a set scope and sequence that must be followed. While the adaptive assessment assigns a set of objectives for a student to work on, the student can work on these lessons in any order they choose, and teachers can assign additional objectives that were not initially assigned through the adaptive assessment.

[18] Experimental estimates of AM's treatment effect on independent, nationally-normed assessments have shown no statistically significant evidence that AM enhances math achievement. Ysseldyke and Bolt (2007) randomly assign elementary and middle school classes to receive access to the Accelerated Math curriculum. They find that treatment classes do not outperform control classes in terms of math achievement on the TerraNova, a popular nationally-normed assessment. Lambert and Algozzine (2009) also randomly assign classes of students to receive access to the AM curriculum to generate causal estimates of the impact of the program on math achievement in elementary and middle school classrooms (N=36 elementary classrooms, N=46 middle school classrooms, divided evenly between treatment and control). Lambert and Algozzine do not find any statistically significant differences between treatment and control students in math achievement as measured by the TerraNova assessment. Nunnery and Ross (2007) use a quasi-experimental design to compare student performance in nine Texas elementary schools and two Texas middle schools who implemented the full School Renaissance Program (including Accelerated Math) to nine comparison schools designated by the Texas Education Agency as demographically similar. Once the study's results were adjusted to account for clustering, Nunnery and Ross's (2007) analysis reveals no statistically significant evidence of improved math performance for elementary or middle school students.

C. Teachers

Fifth grade math teachers at treatment schools received $6 for each academic conference held with a parent in addition to being eligible for monetary bonuses through the HISD ASPIRE program, which rewards teachers and principals for improved student achievement. Each treatment school also appointed a Math Stars coordinator responsible for collecting parent/teacher conference verification forms and organizing the distribution of student reward certificates, among other duties. Coordinators received an individual stipend of $500, which was not tied to performance.

Over the length of the program the average student received $226.67 with a total of $393,038 distributed to students. The average parent received $236.68 with a total of $430,986 distributed to parents. The average teacher received $1,116.48 with a total of $51,358 distributed to teachers. Incentives payments totaled $875,382.

4 Data, Research Design, and Econometric Model

A. Data

We collected both administrative and survey data from treatment and control schools. The administrative data includes first and last name, birth date, address, race, gender, free lunch eligibility, behavioral incidents, attendance, special education status, limited English proficiency (LEP) status, and four measures of student achievement: TAKS math and ELA and STAAR math and reading assessments. Toward the end of the treatment year, the TAKS assessments were administered between April 12 and April 23, 2011, with a retake administered from May 23 to May 25, 2011. At the end of the follwing year, the STAAR assessments were administered from April 24 to April 25, 2012. We use administrative data from 2008-09 and 2009-10 (pre-treatment) to construct baseline controls with 2010-11and 2011-12 (post-treatment) data for outcome measures.

Our main outcome variables are the direct outcomes that we provided incentives for: mastering math objectives via Accelerated Math and attending parent-teacher conferences.

We also examine indirect outcomes that were not directly incentivized, including TAKS math and ELA scale scores, Stanford 10 math and ELA scale scores, and several survey outcomes.

We use a parsimonious set of controls to aid in precision and to correct for any potential imbalance between treatment and control. The most important controls are reading and math achievement test scores from the previous two years and their squares, which we include in all regressions. Previous years' test scores are available for most students who were in the district in previous years (see Table 3 for exact percentages of experimental group students with valid test scores from previous years). We also include an indicator variable that takes on the value of one if a student is missing a test score from a previous year and zero otherwise.

Other individual-level controls include a mutually exclusive and collectively exhaustive set of race dummies pulled from each school district's administrative files, indicators for free lunch eligibility, special education status, and whether a student demonstrates limited English proficiency.[19] Special education and LEP status are determined by HISD Special Education Services and the HISD Language Proficiency Assessment Committee.

We also construct three school-level control variables: percent of student body that is black, percent Hispanic, and percent free lunch eligible. For school-level variables, we construct demographic variables for every 5th grade student in the district enrollment file in the experimental year and then take the mean value of these variables for each school. We assign each student who was present in an experimental school before October 1 to the first school they are registered with in the Accelerated Math database. Outside the experimental group, we assign each student to the first school they attend according to the HISD attendance files, since we are unable to determine exactly when they begin attending school in HISD. We construct the school-level variables based on these school assignments.

[19] A student is income-eligible for free lunch if her family income is below 130 percent of the federal poverty guidelines, or categorically eligible if (1) the student's household receives assistance under the Food Stamp Program, the Food Distribution Program on Indian Reservations (FDPIR), or the Temporary Assistance for Needy Families Program (TANF); (2) the student was enrolled in Head Start on the basis of meeting that program's low-income criteria; (3) the student is homeless; (4) the student is a migrant child; or (5) the student is a runaway child receiving assistance from a program under the Runaway and Homeless Youth Act and is identified by the local educational liaison.

To supplement each district's administrative data, we administered a survey to all parents and students in treatment and control schools.[20] The data from the student survey includes information about time use, spending habits, parental involvement, attitudes toward learning, perceptions about the value of education, behavior in school, and Ryan's (1982) Intrinsic Motivation Inventory. The parent survey includes basic demographics such as parental education and family structure as well as questions about time use, parental involvement, and expectations.

To aid in survey administration, incentives were offered at the teacher level for percentages of student and parent surveys completed. Teachers in treatment and control schools were eligible to receive rewards according to the number of students they taught: teachers with between 1-20 students could earn $250, while teachers with 100 or more students could earn $500 (with fifty dollar gradations in between). Teachers only received their rewards if at least ninety percent of the student surveys and at least seventy-five percent of parent surveys were completed.

In all, 93.4 percent of student surveys and 82.8 percent of parent surveys were returned in treatment schools; 83.4 percent of student surveys and 63.3 percent of parents surveys were returned in control schools. These response rates are relatively high compared to response rates in similar survey administrations in urban environments (Parks et al. 2003, Guite et al. 2006, Fryer 2010).

Table 3 provides descriptive statistics of all HISD 5th grade students as well as those in our experimental group, subdivided into treatment and control. The first column provides the mean, standard deviation, and number of observations for each variable used in our analysis for all HISD 5th grade students. The second column provides the mean, standard deviation, and number of observations for the same set of variables for treatment schools. The third column provides identical data for control schools. The fourth column displays the p-values from a t-test of whether treatment and control means are statistically equivalent. See Online Appendix C for details on how each variable was constructed.

[20] Parent surveys were available in English and Spanish.

Within the experimental group, treatment and control students are fairly balanced, although treatment schools have more black students and fewer white, Asian, LEP, and gifted and talented students. Treatment schools also have lower previous year scores in TAKS math. A joint significance test yields a p-value of 0.643, suggesting that the randomization is collectively balanced along the observable dimensions we consider.

To complement these data, Appendix Figure 2 shows the geographic distribution of treatment and control schools, as well as census tract poverty rates. These maps confirm that our treatment and control schools are similarly distributed across space and are more likely to be in higher poverty areas of a city.

B. Research Design

We use a matched-pair randomization procedure similar to those recommended by Imai et al. (2009) and Greevy et al. (2004) to partition the set of interested schools into treatment and control.[21] Recall that we invited seventy-one schools to sign up for the randomization. Fifty-nine schools chose to sign up. To conserve costs, we eliminated the nine schools with the largest enrollment among the 59 eligible schools that were interested in participating, leaving 50 schools from which to construct 25 matched pairs.

To increase the likelihood that our control and treatment groups were balanced on a variable that was correlated with our outcomes of interest, we used past standardized test scores to construct our matched pairs. First, we ordered the full set of 50 schools by the sum of their mean reading and math test scores in the previous year. Then we designated every two schools from this ordered list as a "matched pair" and randomly drew one member of the matched pair into the treatment group and one into the control group.

C. Econometric model

[21] There is an active debate on which randomization procedures have the best properties. Imbens (2011) summarizes a series of claims made in the literature and shows that both stratified randomization and matched-pairs can increase power in small samples. Simulation evidence presented in Bruhn and McKenzie (2009) supports these findings, though for large samples there is little gain from different methods of randomization over a pure single draw. Imai et al. (2009) derive properties of matched-pair cluster randomization estimators and demonstrate large efficiency gains relative to pure simple cluster randomization.

To estimate the causal impact of providing financial incentives on outcomes, we estimate Intent-To-Treat (ITT) effects, i.e., differences between treatment and control group means. Let Z_s be an indicator for assignment to treatment, let X_i be a vector of baseline covariates measured at the individual level, and let X_s denote school-level variables; X_i and X_s comprise our parsimonious set of controls. Moreover, let ϕ_m denote a mutually exclusive and collectively exhaustive set of matched pair indicators. The ITT effect, π, is estimated from the equation below:

$$achievement_{i,m} = \alpha + X_i\beta + X_s\gamma + Z_s\pi + \phi_m\theta + \varepsilon_{i,m} \tag{11}$$

The ITT is an average of the causal effects for students in schools that were randomly selected for treatment at the beginning of the year and students in schools that signed up for treatment but were not chosen. In other words, ITT provides an estimate of the impact of being *offered* a chance to participate in the experiment. All student mobility between schools after random assignment is ignored. We only include students who were in treatment and control schools as of October 1 in the year of treatment.[22] In HISD, school began August 23, 2010; the first student payments were distributed October 20, 2010.

5 Analysis

5.1 Direct Outcomes

Panels A and B of Table 4 include ITT estimates of treatment effects on incentivized outcomes – AM objectives mastered and parent-teacher conferences attended. Objectives mastered are measured in σ units. Results with and without our parsimonious set of controls are presented in columns (1) and (2), respectively. In all cases, we include matched pair fixed effects. Standard errors are in parenthesis below each estimate. To streamline the pre-

[22] This is due to a limitation of the attendance data files provided by HISD. Accelerated Math registration data confirms students who were present in experimental schools from the beginning of treatment. Using first school attended from the HISD attendance files or October 1 school does not alter the results.

sentation of the experimental results, we focus the discussion in the text on the regressions which include our parsimonious set of controls. All qualitative results are the same in the regressions without controls.

The impact of the financial incentive treatment is statistically significant across all of the direct outcomes we explore. The ITT estimate of the effect of incentives on objectives mastered in AM is 1.087σ (0.031). Treatment parents attended 1.578 (0.099) more parent conferences. Put differently, our aligned incentive scheme caused a 125% increase in the number of AM objectives mastered and an 87% increase in the number of parent-teacher conferences attended in treatment versus control schools.[23]

In addition, we were able to calculate the price elasticity of demand for math objectives by examining the change in AM objectives mastered before and after two unexpected price shocks as seen in Figure 1. After five months of rewarding math objective mastery at a rate of $2 per objective, we (without prompt or advance warning) raised the reward for an objective mastered in AM to $4 for four weeks starting in mid-February and then from $2 to $6 for one week at the beginning of May. Treatment students responded by increasing their productivity; the rate of objective mastery increased from 2.05 objectives per week at the price of $2 per objective up to 3.52 objectives per week at $4 per objective, and 5.80 objectives per week at $6 per objective. Taken at face value, this implies a price elasticity of demand of 0.87.

Taken together, the evidence on the number of objectives mastered and parent conferences attended in treatment versus control schools as well as the response to unexpected price shocks implies that our incentive scheme significantly influenced student and parent behavior.

[23]The average control school actively mastered objectives during 8.16 of 9 payment periods. One school never began implementing the program and six stopped utilizing the program at some point during the year. Of these six, one ceased active use during February, four stopped during March, and one stopped during April. All twenty-five treatment schools actively mastered objectives throughout the duration of the program.

5.2 Indirect Outcomes

In this section, we investigate a series of indirect outcomes – standardized test scores, student investment, parental involvement, attendance, and intrinsic motivation – that are correlated with the outcomes for which we provided incentives. Theoretically, due to misalignment, moral hazard, or psychological factors, the effects of our incentive scheme on this set of outcomes is ambiguous. For these, and other reasons, Kerr (1975) notoriously referred to investigating impacts on indirect outcomes as "the folly of rewarding A, while hoping for B." Still, given the correlation between outcomes such as standardized test scores and income, health, and the likelihood of incarceration, they may be more aligned with the outcomes of ultimate interest than our direct outcomes (Fryer 2011b).

A. Student Test Scores

Panel A of Table 4 presents estimates of the effect of incentives on testing outcomes for which students were not given incentives. These outcomes include Texas' state-mandated standardized test (TAKS). The math and ELA assessments are normalized to have a mean of zero and a standard deviation of one across the city sample. Estimates without and with our parsimonious set of controls are presented in columns (1) and (2), respectively. As before, standard errors are in parentheses below each estimate.

ITT estimates reveal that treatment students outperform control students by 0.081σ (.025) in TAKS math and underperform in TAKS ELA by 0.089σ (.027).[24]

[24] It may be surprising that the impact on math scores is not larger, given the increase in effort on mastering math objectives that were correlated with the Texas state test. One potential explanation is that the objectives in AM are not aligned with those assessed on TAKS. Using Accelerated Math's alignment map, we found that of the 152 objectives in the AM Texas 5th grade library, only 105 (69.1 percent) align with any Texas state math standards (TEKS). Texas state standard alignments are available at `http://www.renlearn.com/fundingcenter/statestandardalignments/texas.aspx` Furthermore, matching the AM curriculum to Texas Essential Knowledge and Skills (TEKS) standards in the six sections of the TAKS math assessment reveals the AM curriculum to be heavily unbalanced; 91 out of the 105 items are aligned with only 3 sections of the TAKS assessment (1, 4, and 6). The treatment effect on the aligned sections is modest in size and statistically significant, 0.137σ (.028). The treatment effect on the remaining (non-aligned) portions of the test is small and statistically insignificant, 0.026σ (.030). Not shown in tabular form. Another, non-competing, explanation is that students substituted effort from another activity that was important for increasing test scores (i.e. paying attention in class) to mastering math objectives.

B. Student and Parent Engagement

The survey results reported in Panel B of Table 4 report measures of student and parent engagement. Students were asked a variety of survey questions including "Did your parents check whether you had done your homework more this year or last year?" and "What subject do you like more, math or reading?" Parents were also asked a variety of questions including "Do you ask your 5th grade student more often about how he/she is doing in Math class or Reading class?" Answers to these questions are coded as binary measures and treatment effects are reported as a percentage change. Details on variable construction from survey responses are outlined in Online Appendix C.

Treatment parents were 7.2 (2.7) percentage points more likely, relative to the control mean of 31 percent, to report that they checked their student's homework more during the treatment year than in the pre-treatment year. Moreover, the increased parental investment was skewed heavily towards math. Treatment parents were 12.2 (2.8) percentage points more likely to ask more about math than reading homework, and treated students were 11.2 (2.3) percentage points more likely to report a preference for math over reading.

C. Attendance and Intrinsic Motivation

The first row of Panel C in Table 4 reports results for student attendance – a proxy for effort. The treatment effect on attendance rates are 0.051σ (0.027) higher than their control counterparts. This amounts to treatment students attending roughly one half of an extra day of school per year.

One of the major criticisms of the use of incentives to boost student achievement is that the incentives may destroy a student's "love of learning." In other words, providing extrinsic rewards can crowd out intrinsic motivation in some situations. There is a debate in social psychology on this issue – see Cameron and Pierce (1994) for a meta-analysis.

To measure the impact of our incentive experiments on intrinsic motivation, we administered the Intrinsic Motivation Inventory, developed by Ryan (1982), to students in our experimental groups.[25] The instrument assesses participants' interest/enjoyment, perceived

[25]The inventory has been used in several experiments related to intrinsic motivation and self-regulation

competence, effort, value/usefulness, pressure and tension, and perceived choice while performing a given activity. There is a subscale score for each of those six categories. We only include the interest/enjoyment subscale in our surveys, as it is considered the self-report measure of intrinsic motivation. To get an overall intrinsic motivation score, we sum the values for these statements (reversing the sign on statements where stronger responses indicate less intrinsic motivation). Only students with valid responses to all statements are included in our analysis of the overall score, as non-response may be confused with low intrinsic motivation.

The final row of Table 4 provides estimates of the impact of our incentive program on the overall intrinsic motivation score of students in our experimental group. The ITT effect of incentives on intrinsic motivation is almost exactly zero – 0.005σ (0.06).

5.3 Heterogenous Treatment Effects

Table 5 investigates treatment effects on number of objectives mastered and state test scores for a set of predetermined subsamples – gender, race/ethnicity, previous year's test score, and whether a student is eligible for free or reduced price lunch.[26]

All regressions include our parsimonious set of controls. Gender is divided into two categories and race/ethnicity is divided into five categories: non-Hispanic white, non-Hispanic black, Hispanic, non-Hispanic Asian and non-Hispanic other race. We only include a racial/ethnic category in our analysis if there are at least one hundred students from that racial/ethnic category in our experimental group; only black and Hispanic subgroups meet this criteria. Eligibility for free lunch is used as an income proxy. We also partition students into quintiles according to their baseline TAKS math scores and report treatment effects for the top and bottom quintiles.

The treatment effect on objectives mastered is statistically larger for girls (1.159σ) than for boys (1.012σ). Hispanic students made the strongest gains on math tests. They also

[e.g., Ryan, Koestner, and Deci (1991) and Deci et al. (1994)].

[26]All other outcomes are in Appendix Tables 3.

mastered more objectives while their parents attended fewer conferences. Students eligible for free lunch showed statistically larger and statistically significant gains on TAKS Math (0.144σ). They also lost less ground in reading; however, the inter-group differences are only marginally significant in reading.

The most noticeable and robust differences occur when we stratify on previous year test scores. Consistent with Proposition 4 from Section 2, high-ability students gain most from the experiment, both in comparison to high-ability students in control schools or low-ability students in treatment schools. For instance, high-ability students master 1.66σ (.117) more objectives, have parents who attend two more parent-teacher conferences, have 0.228σ (.082) higher standardized math test scores and equal reading scores relative to high-ability students in control schools. Conversely, low-ability students master 0.686σ (0.047) more objectives, but score 0.163σ (0.063) *lower* in reading and have similar math test scores compared with low-ability students in control schools.

5.4 Post-Treatment Outcomes

The treatment ended with a final payment to students in June of 2011. A full year after the experiment, we collected data on post-treatment test scores; math and reading tests for treatment and control students during late spring of their sixth grade year.

Recall that in the model, low-ability and high-ability students who are induced to put forth additional effort on a given task learn their type when they observe the results of their additional exertion of effort and that high-ability agents have lower cost of displaced effort. If agents base future effort on their beliefs about their ability-type and update their beliefs in this way, the provision of incentives could lead low-ability agents to exert less effort in the future, while high-ability agents increase their expected return to effort uopn learning they are a high-ability agent and exert more effort in the future.

Table 6 examines lasting treatment effects on standardized test scores and attendance in the year following treatment. Column 1 displays the treatment effects that persisted one full year after all financial incentives were withdrawn for the full group of students with valid

2011-12 test scores. Columns 2 and 3 display the same results for the subgroups of students in the top and bottom quintiles of pre-treatment math test scores. In columns 5 and 6, we restricted our sample to treatment students only and regressed year 1 state test scores on objectives mastered (a measure of effort exerted in math) and predicted the residuals for each student. These residuals capture the difference between a student's expected score on the state test (based upon effort, as measured by objectives mastered) and her actual score. Students were divided into quintiles based upon the size of this residual, with students whose residual is the most negative in the bottom quintile or, "bad shock" group and students with the largest residuals in the top quintile or, "good shock" group. Columns 5 and 6 report the coefficient on a dummy for being in the top or bottom quintile in a regression of second year test scores on residual quintiles and our standard set of parsimonious controls, including two years of lagged test scores. Point estimates are relative to the median quintile, which is omitted from the regression.

While post-treatment effects in the full group sample are statistically insignificant in math, negative effects linger in reading – the effects are -0.042σ (.029) in math and -0.084σ (.030) in reading – the subgroups reveal stark differences between higher and lower achieving students, as well as differences based upon what students may have learned about their ability from their first year effort and resulting test scores. The negative effect on the reading scores of lower-achieving students persist, as lower-achieving treatment students score 0.168σ (.079) lower than lower-achieving control students, and there are significant spillovers into math achievement, where lower-achieving treatment students are outperformed by 0.223σ (.056). Conversely, higher-achieving treatment students outperform their control group peers by 0.134σ (.078) in math and 0.097σ (.086) in reading. While the latter result is statistically insignificant, the positive point estimates in both subjects suggests that any spillover effects were positive or neutral. Furthermore, consistent with the model, we observe that students within the treatment group who experience a "bad shock" in the sense that they underperform on the on the 2010-11 state math test relative to the amount of effort they exerted in AM perform far worse on their 2011-12 standardized tests than students who experience “good shock” in their 2010-11 state math test scores relative to the amount of effort they exerted in AM. Students who experience “bad shocks” score 0.252σ (0.055) lower than students whose test scores are best predicted by their effort in AM, while students who experience “good shocks” score 0.498σ (0.061) higher– a stark difference of 0.75σ between receiving a“bad shock” versus a “good shock” in 2010-11 on students’ 2011-12 test scores.

6 Conclusion

Individuals, even school children, respond to incentives. How we design those incentives to illicit desirable short and longer term responses is far less clear. We demonstrates these complexities with theory and a field experiment.

Our model has two periods and two tasks. In each period, a risk-neutral principal offers a take-it-or-leave-it linear incentive contract to an agent, who, upon accepting the contract, takes two non-verifiable actions which we label “effort.” Effort generates a benefit to the principal and is related to an observable (and contractable) performance measure. We assume that an agent’s type augments their effort in producing output: higher type agents have higher returns to effort than lower type agents, all else equal. The key assumptions are that individuals do not know the production function (e.g. the mapping from input to output) and, upon observing a contract, do not update their beliefs about the production function.[27] It yields four predictions: First, incentives for a given task lead to an increase in effort on that task. Second, incentives for a given task lead to a decrease in effort on the non-incentivized task. Further, the decrease in effort on the non-incentivized task can be more or less for higher-type agents relative to lower-type agents, depending on how substitutable those tasks are in the cost of effort function. The most distinguishing theoretical result concerns the persistent effects of changes in incentives due to agents updating about their ability types. We show that when the agent’s true ability on a given task is sufficiently low, the learning that comes from the provision of incentives is detrimental to the principal. In the absence

[27]As indicated in the introduction, this is a stark assumption for which we provide two justifications.

of incentives the agent would exert some baseline level of effort due to intrinsic motivation and hence learn "little" about her ability. Providing incentives induces more effort than this and hence more learning about their ability type. When agents discover that they are lower-ability than they previously believed, they exert lower effort in period two for any tasks on which there is a positive incentive slope (as in the case of optimal incentives). The average impact of an incentive contract depends on the distribution of across types, among other things.

To better understand these predictions in a real-world laboratory, we analyze new data from a field experiment conducted in fifty traditionally low-performing public schools in Houston, Texas during the 2010-2011 school year. We provided financial incentives to students, their parents, and their teachers for fifth graders in twenty-five treatment schools. Students received \$2 per math objective mastered in Accelerated Math (AM). Parents also received \$2 for each objective their child mastered and \$20 per parent-teacher conference attended to discuss their student's math performance. Teachers earned \$6 for each parent-teacher conference held and up to \$10,100 in performance bonuses for student achievement on standardized tests.

We argue that the data from the field experiment are consistent with the model. Higher-achieving students master 1.66σ more objectives, have parents who attend two more parent-teacher conferences, have 0.228σ higher standardized math test scores and equal reading scores relative to high-achieving students in non-treated schools. Conversely, lower-achieving students master 0.686σ more objectives, have parents who attend 1.5 more parent-teacher conferences, have equal math test scores and 0.163σ lower reading scores. Put differently, higher-achieving students put in significant effort and were rewarded for that effort in math without a deleterious impact in reading. Lower-achieving students also increased effort on the incentivized task, but did not increase their math scores and their reading scores decreased significantly.

Consistent with the fourth – and most distinguishing– prediction of the model, higher-achieving students continue to do well, maintaining a positive treatment effect in math and a zero effect in reading, one year after the incentives are taken away. Lower-achieving students, however, exhibit large and statistically significant decreases in both math [-.223σ (0.056)] and reading achievement [-.168σ (0.079)] after the incentives are removed. We argue that this is most likely explained by students learning about their own ability and not decreases in intrinsic motivation. The treatment effect on the latter, gleaned from survey data, is small and statistically insignificant.

Taken together, both the theoretical model and the experimental results offer a strong cautionary tale on the use of financial incentives when individuals may not know the production function.

References

[1] Agarwal, Vikas, Naveen D. Daniel, and Narayan Y. Naik. 2009. "Role of Managerial Incentives and Discretion in Hedge Fund Performance." Journal of Finance. 64(5): 2221-2256.

[2] Angrist, Joshua D., Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." The American Economic Review, 92(5): 1535-1558.

[3] Angrist, Joshua D., Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." The American Economic Review, 96(3): 847-862.

[4] Angrist, Joshua D., Daniel Lang, and Philip Oreopoulos. 2006. "Lead Them to Water and Pay Them to Drink: An Experiment with Services and Incentives for College Achievement." NBER Working Paper No. 12790.

[5] Angrist, Josh D., and Victor Lavy. 2009. "The Effect of High-Stakes High School

Achievement Awards: Evidence from a Group-Randomized Trial." American Economic Review, 99(4): 1384-1414.

[6] Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2008. "Conditional Cash Transfers in Education: Design Features, Peer and Sibling Effects: Evidence from Randomized Experiment in Colombia." NBER Working Paper No. 13890.

[7] Behrman, Jere R., Piyali Sengupta and Petra Todd. 2005. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." Economic Development and Cultural Change, 54: 237-275.

[8] Bettinger, Eric. 2010. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." NBER Working Paper No. 16333.

[9] Bruhn, Miriam and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." American Economic Journal: Applied Economics, 1(4): 200-232.

[10] Bucklin, Barbara R. and Alyce M. Dickinson. 2001. "Individual Monetary Incentives: A Review of Different Types of Arrangements Between Performance and Pay." Journal of Organziational Behavior Management, 21(3): 45-137.

[11] Cameron, Judy and W. David Pierce. 1994. "Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis." Review of Educational Research, 64(3): 363-423.

[12] Condly, Steven J., Richard E. Clark, and Harold D. Stolovich. 2003. "The Effects of Incentives on Workplace Performance: A Meta-Analytic Review of Research Studies. Performance Improvement: 16(3) 46-63.

[13] Deci, Edward L. 1972. "The Effects of Contingent and Noncontingent Rewards and Controls on Intrinsic Motivation." Organizational Behavior and Human Performance, 8:

217-229.

[14] Deci, Edward L. 1975. Intrinsic Motivation. New York: Plenum.

[15] Deci, Edward L., Haleh Eghrari, Brian C. Patrick and Dean R. Leone. 1994. "Facilitating Internalization: The Self-Determination Theory Perspective." Journal of Personality, 62(1): 119-142.

[16] Duflo, Esther, and Rema Hanna (forthcoming). "Incentives Work: Getting Teachers to Come to School." American Economic Review.

[17] Fryer, Roland G. 2010. "Financial Incentives and Student Achievement: Evidence From Randomized Trials." NBER Working Paper No. 15898.

[18] Fryer, Roland G. 2011a. "Financial Incentives and Student Achievement: Evidence From Randomized Trials." Quarterly Journal of Economics. 126 (4).

[19] Fryer, Roland G. 2011b. "Racial Inequality in the 21st Century: The Declining Significance of Discrimination." Forthcoming in *Handbook of Labor Economics, Volume 4*, Orley Ashenfelter and David Card eds.

[20] Fryer, Roland G. (forthcoming) "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." Journal of Labor Economics, forthcoming.

[21] Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." American Economic Journal: Applied Economics, 2(3): 205-227.

[22] Gneezy, Uri and Aldo Rustichini. 2000. "Pay Enough or Don't Pay at All." The Quarterly Journal of Economics, 115(3): 791-810.

[23] Greevy, Robert, Bo Lu, and Jeffrey H. Silber. 2004. "Optimal multivariate matching before randomization." Biostatistics 5: 263-275.

[24] Guite, H., C. Clark, and G. Ackrill. 2006. The Impact of Physical and Urban Environment on Mental Well-Being. Public Health, 120(12): 1117-1126.

[25] Hahn, A., T. Leavitt, and P. Aaron. 1994. "Evaluation of the Quantum Opportunities Program (QOP). Did the program work? A report on the post secondary outcomes and cost-effectiveness of the QOP program (1989-1993)." Massachusetts. (ERIC Document Reproduction Service No. ED 385 621).

[26] Hanushek, Eric A. 2007. "Education Production Functions: Developed Country Evidence," in *International Encyclopedia of Education, Third Edition*, Penelope Peterson, Eva Baker, and Barry McGaw (eds.)

[27] Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." Journal of Law, Economics, and Organization. 7: 24-52.

[28] Imai, Kosuke, Gary King, and Clayton Nall. 2009. "The Essential Role of Pair Matching in Cluster Randomized Experiments." Statistical Science 24(1): 29-53.

[29] Imbens, Guido. 2011. Experimental Design for Unit and Cluster Randomized Trials. Conference Paper, International Initiative for Impact Evaluation.

[30] Kerr, Steven. 1975. "On the Folly of Rewarding A, While Hoping for B." The Academy of Management Journal, 18(4): 769-783.

[31] Kohn, Alfie. 1993. Punished by Rewards. Boston: Houghton Mifflin Company.

[32] Kohn, Alfie. 1996. "By All Available Means: Cameron and Pierce's Defense of Extrinsic Motivators." Review of Educational Research, 66(1): 1-4.

[33] Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." Review of Economics and Statistics. 91(3): 437-456.

[34] Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." The Quarterly Journal of Economics, 114(2): 497-532.

[35] Lazear, Edward P. 2000. "Performance Pay and Productivity." American Economic Review. 90(5): 1346-1361.

[36] Lazear, Edward P. 2001. "Educational Production." Quarterly Journal of Economics. 96(3): 777-803.

[37] Ledford, Gerald E., Edward E. Lawler III, and Susan A. Mohrman. 1995. "Reward Innovations in Fortune 1000 Companies." Compensation Benefits Review. 27(4): 76-80.

[38] Mitra, Atul, Nina Gupta, and G. Douglas Jenkins, Jr. 1995. "The Case of the Invisible Merit Raise: How People See Their Pay Raises." Compensation Benefits Review. 27(3): 71-76.

[39] Muralidharan, Karthik and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." Journal of Political Economy, 119: 39-7.

[40] Murphy, Kevin J. 1998 "Executive Pay," in *Handbook of Labor Economics, Vol. 3*, Orley Ashenfelter and David Card (eds.).

[41] Oosterbeek, Hessel, Edwin Leuven, and Bas van der Klaauw. 2010. "The Effect of Financial Rewards on Students' Acheivement: Evidence From a Randomized Experiment." Journal of the European Economic Association. 8(6): 1243-1265.

[42] Paarsch, Harry J. and Bruce Shearer. "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records." 2000. International Economic Review. 41(1): 59-92.

[43] Parks, S. E., R. A. Housemann, and R. C. Brownson. 2003.Differential Correlates of Physical Activity in Urban and Rural Adults of Various Socioeconomic Backgrounds

in the United States. Journal of Epidemiology and Community Health, 57(1): 29-35

[44] Ryan, Richard M. 1982. "Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory." Journal of Personality and Social Psychology, 63: 397-427.

[45] Ryan, Richard M., Richard Koestner, and Edward L. Deci. 1991. "Ego-Involved Persistence: When Free-Choice Behavior is Not Intrinsically Motivated." Motivation and Emotion, 15(3): 185-205.

[46] Smiley, Patricia A. and Carol S. Dweck. "Individual Differences in Achievement Goals among Young Children." Child Development. 65(6): 1723-1743.

[47] Springer, Matthew G., Dave Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. 2010. "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching." Conference Paper, National Cneter on Performance Incentives.

[48] Todd, Petra E. and Kenneth I. Wolpin. "On the Specification and Estimation of the Production Function for Cognitive Achievement." The Economic Journal. 113: F3-F33.

[49] Wagner, Barry M. and Deborah A. Phillips. 1992. "Beyond Beliefs: Parent and Child Behaviors and Children's Perceived Academic Competence." Child Development. 63(6): 1380-1391.

[50] Wiatrowski, William J. 2009. "The Effect of Incentive Pay on Rates of Change in Wages and Salaries." U.S. Bureau of Labor Statistics, Compensation and Working Conditions Online. http://www.bls.gov/opub/cwc/cm20091120ch01.htm

7 Online Appendix A: Proofs of Propositions (Not For Publication)

Proof of Proposition 1. From equation (8) apply the Milgrom-Shannon monotonicity theorem to conclude that the objective function has decreasing differences (b_1^*, σ_1^2) and hence that $db_1^*/d\sigma_1^2 < 0$ and that it has increasing difference in (b_1^*, σ_2^2) and hence that $db_1^*/d\sigma_2^2 > 0$. ■

Proof of Proposition 2. For parts (a) and (b): from equation (6) note that $de_1^*/db_1 > 0$ and that $de_2^*/db_1 < 0$ and analogously for e_2^*. Part (c): note that

$$\frac{\partial^2 e_2}{\partial b_1 \overline{\alpha}_1} = \frac{\delta}{\delta^2 - 1}, \tag{12}$$

is of indeterminate sign. ■

Proof of Proposition 3. Trivial. Follows directly from the intermediate value theorem. ■

Proof of Proposition 4. Consider increasing $b_{1,1}$ as stated in the proposition. Recall that the agent's posterior belief after observing period 1 output is

$$E[\alpha_i|m_i^1] = \bar{\alpha}_i \left(\frac{\sigma_i^2}{\mu_i^2 + \sigma_i^2}\right) + \frac{2(\alpha_i - \overline{\alpha}_i)b_1 - 2(\alpha_{-i} - \overline{\alpha}_{-i})b_2\delta + b_2^2\delta\eta\mu_2^2 - b_1^2\eta\mu_1^2}{2\,(1-\delta^2)} \left(\frac{\mu_i^2}{\mu_i^2 + \sigma_i^2}\right).$$

Consider two agents 1 and 2 with $\alpha_1^1 > \alpha_2^1$. The difference in posterior beliefs is $E[\alpha_1^1|m_i^1] - E[\alpha_1^2|m_i^1]$. Since they have a common prior the term, and thus the difference is

$$\frac{2b_1(\alpha_1^1 - \alpha_1^2) - 2\delta b_2(\alpha_2^1 - \alpha_2^2)}{2(1-\delta^2)} \left(\frac{\mu_i^2}{\mu_i^2 + \sigma_i^2}\right).$$

Note that since $\alpha_1^1 > \alpha_1^2$ by construction, this has increasing differences in (b_1, α_1). By the definition of conditional probability it must be that $E[E[\alpha|m_1]] = E[\alpha] = \overline{\alpha} > 0$. Since $E[\alpha_1|m_i^1] - E[\alpha_1|m_i^1] = 0$ for any α and appealing to the intermediate value theorem the result is established. ■

8 Online Appendix B: Implementation Manual (Not For Publication)

Schools

We identified 71 low-performing elementary schools in the district based upon the average grade 5 scores on the Texas Assessment of Knowledge and Skills (TASKS) that would benefit from inclusion in the Math Stars incentive program. On Thursday, September 2, 2010, HISD leadership held an introductory meeting with principals and math teachers from these low-performing elementary schools. After presenting an overview of the research design we invited them to commit to participate by signing a pledge to implement the Math Stars program with fidelity to the research design.

Schools had five days to consider their commitment to the program (within a day, however, over two-thirds of the schools invited had already indicated their commitment and interest by signing a School Commitment Letter.) By Tuesday, September 7, 60 schools had elected to participate in the random selection process, and we conducted a random lottery to select the 25 treatment schools and the 25 control schools.

Students

HISD decided that students and parents at selected schools would be automatically enrolled in the program. Parents could choose not to participate and return a signed opt-out form at any point during the school year. HISD further decided that students and parents were required to participate jointly: students could not participate without their parents and vice versa.

Software and Incentive Structure

The Accelerated Math platform creates math assignments tailored to each student's ability level, enabling students to take brief online assessments to gauge achievement in mathematics. For fifth grade, math objectives fall into the following subject areas: Number Sense and Operations; Algebra; Geometry and Measurement; and Data Analysis, Statistics, and Probability.

Students began the program year by taking an initial diagnostic assessment to measure mastery of math concepts, after which AM creates customized practice assignments that focus specifically on areas of weakness. Teachers assign these custom assignments and students are then able to print the assignments and take them home to work on (with or without their parents). Each assignment has six questions, and students must answer at least five questions correctly to receive credit. Students scan their completed assignments into AM, and the assignments are graded electronically. Teachers then administer an AM test that serves as the basis for potential rewards: students are given credit for official mastery by answering at least four out of five questions correctly.

Students: Students earned $2 for every objective mastered. Students who reached the 200 objectives threshold were declared Math Stars and received a $100 completion bonus and special certificate. Additional monetary incentives were introduced during the program: during the sixth pay period (mid-February to mid-March) students received $4 for every objective mastered; during the final week of the eighth pay period (the first week of May), students received $6 for every objective mastered.

Parents: Parents of children at treatment schools earned up to \$160 for attending eight parent-teacher review sessions (\$20/each) in which teachers presented student progress using Accelerated Math Progress Monitoring dashboards. Parents and teachers were both required to sign the student progress dashboards and submit them to their schools Math Stars coordinator in order to receive credit. Additionally, parents earned \$2 for their child's mastery of each AM curriculum objective, as long as they attended at least one conference with their child's teacher. This requirement also applied retroactively: if a parent first attended a conference during the final pay period, the parent would receive a lump sum of \$2 for each objective mastered by their child to date. Parents were not instructed on how to help their children complete math worksheets.

Teachers: Fifth grade math teachers at treatment schools received \$6 for each academic conference held with a parent in addition to being eligible for monetary bonuses through the HISD ASPIRE program, which rewards teachers and principals for improved student achievement. Each treatment school also appointed a Math Stars coordinator responsible for collecting parent/teacher conference verification forms and printing and distributing student reward certificates, among other duties. Each coordinator received a stipend of \$500, but this amount was not tied to performance.

Principals: Principals at treatment schools were eligible for monetary bonuses through the HISD ASPIRE program, which rewards teachers and principals for improved student achievement.

Training and Program Launch

Once schools were selected, the Accelerated Math program was ordered for treatment and control schools, as well as computers and scanners for each school (depending on the number of students and classrooms). AM was installed in treatment schools on September 10 and control schools on September 20. HISD also hired a district-based program manager who was trained in using AM as well as a technology support staff member.

On September 10, a welcome packet in English and Spanish was sent home with students. The packet included a detailed description of the program, a program calendar, answers to frequently asked questions, and an opt-out form. Parents who decided they did not want their student(s) to participate in the incentive component of the Math Stars program were able to return a signed opt-out form at any point during the school year; however, students were not able to opt out of using the Accelerated Math platform.

Meanwhile, treatment schools identified in-school coordinators within one day of being randomly selected; coordinators primary duties included collecting parent-teacher conference sheets and distributing checks and reward certificates to students on pay day. To effectively train participating schools staff to use the Accelerated Math program, Renaissance Learning staff conducted teacher and coordinator training in treatment schools the week beginning September 13 (teachers in control schools were trained from September 28-29.)

Teacher training consisted of coaching teachers in how to use the Accelerated Math platform to provide practice and assessment opportunities for students at different skill levels. To ensure differentiated instruction, students were able to test within multiple grade levels of objectives. Therefore, a library or bank of Accelerated Math objectives, practice questions, and assessments – spanning second through seventh grades – were available from which teachers could pull assignments that students could master. However, starting in February – four full months after the beginning of the program – teachers were restricted from drawing objective assignments from libraries below fourth grade equivalency.

After brief site visits to ensure that experimental schools' technological infrastructures were properly in place, teachers were re-trained in how to use Star Math (a companion program to the Accelerated Math platform that was already in place in the HISD schools), which allows classroom teachers to administer a customized diagnostic test to students to assess skill levels within certain grade-level objectives. Therefore, to determine the grade level at which each student should begin their mastery of objectives, teachers began administering student diagnostic assessments the week beginning Monday, September 20. Within two days, 92 percent of students in treatment schools had taken the diagnostic.

Payment Process

Preparation and Set-up: At the conclusion of each pay period, the district-based program manager would begin processing student and parent payments along two fronts: first, extracting student performance data from the Accelerated Math platform, removing students who opted out, and calculating student rewards ($2/per objective mastered); second, collecting parent-teacher conference dashboards from school coordinators and inputing parent attendance figures. These two data points are consolidated in a pay file and organized by school.

After all parent conference data was collected and inputed, the pay file was sent to EdLabs to complete the payment algorithm and conduct a few basic audits. The pay file was then sent back to the district program manager, who reformatted and finalized the file for the HISD finance office, who uploaded payment information to JP Morgan Chase. Checks were printed, bundled by school, and delivered to each school.

EdLabs also used the pay file to create reward certificates for every student receiving a payment. The certificate detailed how many math objectives the student mastered during the last period, the cumulative total, and the current financial earnings. When students passed the 200 objective threshold, they received a special certificate in addition to their $100 bonus.

Payment Logistics: School coordinators received student and parent checks and student certificates one day prior to pay day. Each school planned pay day differently, but there was striking uniformity: typically a small assembly was held in the cafeteria during which checks and certificates were distributed and students were recognized for their achievements. Parents were often in attendance as well to acknowledge their children and receive their checks.

Bonus Rounds

The first several pay periods of Math Stars yielded high rates of participation among both students (i.e. percentage of students mastering at least one objective and receiving payment) and parents (i.e. percentage of parents attending a conference with their students teacher). As a result of smooth implementation and general enthusiasm about the program among students and staffmembers, HISD and EdLabs introduced two bonus rounds: during the entire sixth pay period, (February 14 through March 11), students received $4 (rather than the usual $2) for each objective mastered. During the final week of the eighth pay period (May 2 through May 5), students received $6 for each objective mastered. These changes were communicated to students primarily through posters hung throughout the school and flyers sent home in weekly folders.

There were two primary objectives in introducing these bonus rounds: first, the additional incentive was meant to strengthen students preparation for end-of-year testing. The first ($4) bonus round took place just prior to the Texas Assessment of Knowledge and Skills (TAKS), while the second ($6) bonus round took place prior to the Stanford 10. Second, a sub-experiment was being conducted to estimate a demand curve for math objectives; i.e. asking whether a student will devote more effort to mastering math objectives relative to the increase in the reward.

Site Visits

In an effort to gather extensive qualitatitve data on the implementation of HISD's Math Stars program, EdLabs conducted brief site visits to all 25 treatment schools.

EdLabs observed classrooms, interviewed students, teachers, and school leaders, and developed, with extensive help from HISD program personnel, a site visit rubric. In addition to providing a comprehensive collection of qualitative school-level data to use in the evaluation of the Math Stars program (i.e. correlating school-level performance with observed implementation indicators), the site visits also supplied the district-based program manager with additional best practices to share with other schools during the last few pay periods of the program.

9 Online Appendix C: Variable Construction (Not For Publication)

Attendance Rates
When calculating the school-level attendance rate, we consider all the presences and absences for students when they are enrolled at each school. Individual attendance rates account for all presences and absences for each particular student, regardless of which school the student had enrolled in when the absence occurred.

Effort Index
To gauge how treatment affected students' effort, we surveyed students about how strongly they agreed with the following six statements: (1) Students in my school are usually on time for class (2) Students in my classes usually turn in their homework (3) Students in my classes usually ask questions (4) I am satistified with what I have achieved in my classes (5) I have pushed myself to completely understand my lessons in school (6) I could do much better in school if I worked harder. In each case, students were instructed to indicate whether they believed the statement is totally untrue, mostly untrue, somewhat true, mostly true, or totally true. This responses were coded on an integer scale ranging from 1-5, with 1 corresponding to "totally untrue." To construct our index of effort, we added up the numeric values on all five responses (inverting the sign on question 6) and normalized the sum to have a mean of zero and a standard deviation of one. We only calculate an index for students with a valid response for all five statements, as nonresponse might otherwise be confused with strong disagreement. When individual questions appear as dependent variables in regressions, they were normalized similarly.

Free Lunch
Regressions include a dummy variable equal to one if a student is eligible for free or reduced-price lunch.

Gifted and Talented
HISD offers two Gifted and Talented initiatives: Vanguard Magnet, which allows advanced students to attend schools with peers of similar ability, and Vanguard Neighborhood, which provides programming for gifted students in their local school. We consider a student gifted if he or she is involved in either of these programs.

Motivation Index
We disseminated part of the Intrinsic Motivation Inventory, developed by Ryan (1982), to students in our experimental group. The instrument contains many modules, but we limited our questions to those in the interest/enjoyment subscale in our surveys as it is considered the self-reported measure of intrinsic motivation. The interest/enjoyment subscale consists of seven statements on the survey: (1) I enjoy doing schoolwork very much; (2) doing schoolwork is fun; (3) I thought this was a boring activity; (4) doing schoolwork does not hold my attention at all; (5) I would describe doing schoolwork as very interesting; (6) I

think doing schoolwork is quite enjoyable; and (7) while I am doing schoolwork, I think about how much I enjoyed it. Respondents are asked how much they agree with each of the above statements on a seven-point Likert scale ranging from "not at all true" to "very true." To get an overall intrinsic motivation score, one adds up the values on each statement (reversing the sign on statements (3) and (4)). Only students with valid responses on each statement are included in our analysis of the overall score, as non-response may be confused with low intrinsic-motivation. When reporting results, we report effects on scores normalized to have a mean of zero and a standard deviation of one.

Special Education and Limited English Proficiency
These statuses are determined by HISD Special Education Services and the HISD Language Proficiency Assessment Committee, respectively; they enter into our regressions as dummy variables. We do not consider students who have recently transitioned out of LEP status to be of limited English proficiency.

Suspensions
The school-level count of suspensions includes both in-school and out-of-school suspensions, regardless of the nature of the infraction.

Race/Ethnicity
We code the race variables such that the five categories – white, black, Hispanic, Asian and other – are collectively exhaustive and mutually exclusive. Hispanic ethnicity is an absorbing state. Hence "white" implies non-Hispanic white, "black" non-Hispanic black, and so on.

Survey Responses
Some of the indirect outcomes reported in the paper include survey responses. We include two questions from the student survey. First, students were asked "Did your parents check your homework this year more than last year?" We code responses of "more this year" as 1 and responses of either "more last year" or "about the same" as 0. Second, students were asked "What subject do you like better, math or reading?" We code responses of "math" as 1 and "reading" as 0.

We also report the results of one question from the parent survey. Parents were asked "Do you ask your 5th grade student more often about how he/she is doing in math class or reading class?" We code responses of "math class" as 1 and responses of either "reading class" or "no difference" as zero.

Teacher Value-Added
HISD officials provided us with 2009-10 value-added data for 3,883 middle and elementary school teachers. In Table 2, we present calculations based on the district-calculated Cumulative Gain Indices. We normalize these indices such that the average teacher in each subject has a score of zero and the sample standard deviation is one. These scores are then averaged within each school.

Test Scores

We observe results from the Texas Assessment of Knowledge and Skills (TAKS) and the Stanford 10. For ease of interpretation, we normalize raw scores to have a mean of zero and a standard deviation of one within grades, subjects, and years.

Treatment

Due to a limitation in the attendance data provided by HISD, we are unable to determine the dates on which students enrolled in their current schools. AM registration files provides a "snapshot" file that records each students' enrolled school as of October 1. We include students in one of the 25 treatment schools on October 1, 2010 in our treatment group (the control group is defined similarly). Our results are not sensitive to changing the treatment assignment based on the first school attended during the 2010-11 school year.

Table 1: Summary of Math Stars Houston Incentives Experiment

Schools	50 (of 70 eligible) HISD schools opted in to participate, 25 schools randomly chosen for treatment. All treatment and control schools were provided complete Accelerated Mathematics software, training, and implementation materials (handouts and practice exercises).
Treatment Group	1,693 5th grade students: 27.5% black, 70.1% Hispanic, 55.5% free lunch eligible
Control Group	1,735 5th grade students: 25.7% black, 68.2% Hispanic, 53.6% free lunch eligible
Outcomes of Interest	TAKS State Assessment, STAAR State Assessment (post-treatment),Number of Math Objectives Mastered, Parent Conference Attendance, Measures of Parent Involvement, Measures of Student Motivation and Effort
Test Dates	Year 1: TAKS: April 12-23, 2011; TAKS Retake: May 23-25, 2011; Stanford 10: May 8-10, 2011 Year 2: STAAR: April 24-25, 2012
Objectives Database	Students took a diagnostic test and were assigned math objectives to practice based upon their measured deficiencies.
Incentive Structure	Students paid $2 per objective to practice a math objective and pass a short test to ensure they mastered it.
Additional Incentives	$100 for mastering 200th objective (cumulatively)
Frequency of Rewards	Paydays were held every 3-4 weeks
Operations	$875,000 distributed in incentives payments, 99% consent rate. 2 dedicated project managers.

Notes. Each row describes an aspect of treatment indicated in the first column. Entries are descriptions of the schools, students, outcomes of interest, testing dates, objectives database, incentive structure, additional incentives, frequency of rewards and operations. See Appendix A for more details. The numbers of treatment and control students given are for those students who have non-missing reading or math test scores.

Table 2: Pre-Treatment Characteristics of Non-Experimental and Experimental Schools

	Non-Exp. 5th Grade	Exp. 5th Grade	E vs. NE p-value	Treatment	Control	T vs. C p-value
Teacher Characteristics						
Percent male	0.161	0.183	0.105	0.174	0.191	0.317
	(0.079)	(0.078)		(0.074)	(0.082)	
Percent black	0.322	0.370	0.307	0.366	0.374	0.777
	(0.255)	(0.292)		(0.330)	(0.257)	
Percent Hispanic	0.343	0.365	0.547	0.352	0.377	0.417
	(0.213)	(0.202)		(0.222)	(0.183)	
Percent white	0.290	0.222	0.033	0.236	0.207	0.668
	(0.233)	(0.158)		(0.141)	(0.176)	
Percent Asian	0.034	0.032	0.798	0.029	0.035	0.315
	(0.039)	(0.032)		(0.030)	(0.035)	
Percent other race	0.010	0.011	0.838	0.015	0.007	0.224
	(0.015)	(0.022)		(0.026)	(0.016)	
Mean teacher salary / 1000	51.942	52.079	0.674	52.088	52.071	0.523
	(2.058)	(1.848)		(1.706)	(2.014)	
Mean years teaching experience	11.878	12.082	0.657	12.222	11.942	0.326
	(2.781)	(2.656)		(2.476)	(2.870)	
Mean Teacher Value Added: Math	0.040	-0.162	0.031	-0.211	-0.113	0.456
	(0.468)	(0.586)		(0.417)	(0.722)	
Mean Teacher Value Added: Reading	0.040	-0.121	0.080	-0.128	-0.113	0.779
	(0.465)	(0.566)		(0.411)	(0.696)	
Student Body Characteristics						
# of suspensions per student	0.096	0.106	0.606	0.087	0.126	0.883
	(0.096)	(0.155)		(0.108)	(0.192)	
# of days suspended per student	0.214	0.261	0.365	0.225	0.297	0.925
	(0.988)	(0.344)		(0.290)	(0.395)	
Total Enrollment 2009-2010	727.467	593.068	0.000	606.522	579.251	0.718
	(202.807)	(142.169)		(163.744)	(117.878)	
Number of Schools	130	50		25	25	

Notes: This table reports school-level summary statistics for our aligned incentives experiment. The non-experimental sample includes all HISD schools with at least one 5th grade class in 2009-10. Column (3) reports p-values on the null hypthesis of equal means in the experimental and non-experimental sample. Column (6) reports the same p-value for treatment and control schools. Each test uses heteroskedasticity-robust standard errors, and the latter test controls for matched-pair fixed effects.

Table 3: Student Pre-Treatment Characteristics

	HISD 5th Grade	Treatment	Control	T vs. C. p-value
Student Characteristics				
Male	0.510	0.526	0.525	0.504
	(0.500)	(0.499)	(0.500)	
White	0.078	0.019	0.046	0.000
	(0.268)	(0.138)	(0.211)	
Black	0.248	0.275	0.257	0.015
	(0.432)	(0.447)	(0.437)	
Hispanic	0.632	0.701	0.682	0.876
	(0.482)	(0.458)	(0.466)	
Asian	0.030	0.001	0.009	0.002
	(0.172)	(0.035)	(0.094)	
Other Race	0.012	0.003	0.006	0.364
	(0.109)	(0.055)	(0.077)	
Special Education Services	0.098	0.108	0.086	0.668
	(0.297)	(0.311)	(0.281)	
Limited English Proficient	0.307	0.293	0.336	0.017
	(0.461)	(0.455)	(0.473)	
Gifted and Talented	0.193	0.138	0.166	0.040
	(0.394)	(0.345)	(0.373)	
Economically Disadvantaged	0.828	0.929	0.909	0.219
	(0.377)	(0.257)	(0.287)	
Free or Reduced Price Lunch	0.513	0.555	0.536	0.349
	(0.500)	(0.497)	(0.499)	
TAKS Math 09-10	0.000	-0.142	-0.082	0.043
	(1.000)	(0.944)	(0.954)	
TAKS ELA 09-10	0.000	-0.166	-0.152	0.629
	(1.000)	(0.934)	(0.956)	
Missing Previous Math Scores	0.129	0.117	0.114	0.448
	(0.336)	(0.321)	(0.317)	
Missing Previous ELA Scores	0.134	0.125	0.122	0.514
	(0.340)	(0.331)	(0.327)	
p-value from joint F-test				0.643
Student Outcomes				
Participated in Program	0.111	0.966	0.001	0.000
	(0.314)	(0.180)	(0.034)	
Periods Treated	0.944	8.473	0.003	0.000
	(2.717)	(1.739)	(0.107)	
Observations	15389	1693	1735	3428

Notes: This table reports summary statistics for our aligned incentives experiment. The sample is restricted to 5th grade students with valid test score data for the 2010 - 2011 school year. Column (4) reports p-values on the null hypothesis of equal means in treatment and control groups using heteroskedasticity-robust standard errors and controls for matched-pair fixed effects.

Table 4 - Mean Effect Sizes (Intent to Treat Estimates): Indirect Outcomes

	Raw	Controlled
A. Student Achievement		
State Math 10-11	0.077***	0.081***
	(0.024)	(0.025)
	3128	3128
State ELA 10-11	-0.084***	-0.077***
	(0.026)	(0.027)
	3108	3108
Aligned State Math 10-11	0.129***	0.137***
	(0.027)	(0.028)
	3090	3090
Unaligned State Math 10-11	0.023	0.026
	(0.029)	(0.030)
	3090	3090
B. Survey Outcomes		
Parents check HW more	0.036	0.071***
	(0.024)	(0.027)
	2041	2041
Student prefers Math to Reading	0.118***	0.112***
	(0.021)	(0.023)
	2356	2356
Parent asks about Math more than Rdg.	0.115***	0.122***
	(0.024)	(0.028)
	1908	1908
Conferences Attended	1.639***	1.572***
	(0.089)	(0.099)
	2052	2052
Objectives Mastered	0.978***	1.087***
	(0.029)	(0.031)
	3292	3292
C. Attendance and Motivation		
Attendance 2010-2011	0.045*	0.050*
	(0.026)	(0.027)
	3187	3187
Intrinsic Motivation	0.041	0.006
	(0.056)	(0.060)
	2004	2004

Notes: This table reports ITT estimates of the effects of our aligned incentives experiment on various test scores and survey responses. Testing and attendance variables are drawn from HISD attendance files and standardized to have a mean of 0 and standard deviation of 1 among 5th graders with valid test scores. The survey responses included here are coded as zero-one variables; The effort and intrinsic motivation indices are constructed from separate survey responses; their construction is outlined in detail in the text of this paper and Online Appendix B. Raw regressions include controls for previous test scores, their squares, and matched-pair fixed effects. Controlled regressions also include controls for the gender, race, free lunch eligibility, special education status, and whether the student spoke English as second language. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 5: Mean Effect Sizes (Intent to Treat) By Subsample

	Whole Sample	*Gender* Male	Female	p-val	*Race* Black	Hispanic	p-val	*Free Lunch* Yes	No	p-val	*Math Quintile* Bottom	Top	p-val
A. Incentivized Outcomes													
Objectives Mastered	1.087***	1.012***	1.159***		0.816***	1.114***		1.096***	1.055***		0.686***	1.660***	
	(0.031)	(0.045)	(0.043)	0.017	(0.045)	(0.045)	0.000	(0.043)	(0.047)	0.519	(0.047)	(0.117)	0.000
	3292	1728	1554		857	2283		1774	1492		694	423	
B. Non-Incentivized Outcomes													
State Math 10-11	0.081***	0.106***	0.040		-0.002	0.104***		0.144***	-0.006		-0.004	0.228***	
	(0.025)	(0.035)	(0.037)	0.183	(0.056)	(0.033)	0.101	(0.034)	(0.037)	0.003	(0.049)	(0.082)	0.011
	3128	1636	1491		828	2165		1687	1421		663	428	
State ELA 10-11	-0.077***	-0.067*	-0.090**		-0.069	-0.076**		-0.033	-0.122***		-0.165***	0.023	
	(0.027)	(0.037)	(0.039)	0.678	(0.071)	(0.033)	0.926	(0.038)	(0.041)	0.106	(0.063)	(0.083)	0.060
	3108	1616	1491		821	2151		1677	1411		659	427	

Notes: This table reports ITT estimates of the effects of the experiment on incentivized and non-incentivized outcomes for a variety of subsamples. All regressions follow the controlled specification described in the notes of previous tables. All test outcomes are standardized to have mean zero and standard deviation one among all HISD fifth graders. *** = significant at 1 percent level, ** = significant at 5 percent level, and * = significant at 10 percent level.

Table 6: Mean Effect Sizes (Intent to Treat) on Second Year Outcomes By Subsample

	Full	Previous Year Math Achievement			Bad	Good	
	Sample	Bottom Quintile	Top Quintile	p-value	Shock	Shock	p-value
State Math 11-12 (Post-treatment)	-0.042	-0.223***	0.134*		-0.252***	0.498***	
	(0.029)	(0.056)	(0.078)	0.000	(0.055)	(0.061)	0.000
	2461	511	332		375	230	
State Reading 11-12 (Post-treatment)	-0.071**	-0.170**	0.103		-0.196***	0.156**	
	(0.029)	(0.080)	(0.086)	0.013	(0.056)	(0.063)	0.000
	2458	516	336		375	230	
Attendance 2011-2012	0.011	0.084	0.018		-0.070	0.040	
	(0.035)	(0.091)	(0.070)	0.538	(0.075)	(0.072)	0.147
	2598	588	342		375	230	

Notes: Columns 1-3 report ITT estimates of the effects of the experiment on year 2 test scores and attendance. Columns 5 and 6 report regression coefficients from a regression of year 2 outcomes on dummies for whether a student received a large negative shock relative to his or her predicted year 1 test score (predicted by objectives mastered in Accelerated Math, a measure of effort). Students are broken into quintiles by the size their residuals from a regression of year 1 test scores on objectives mastered, and students with large negative residuals are in the bottom quintile, having received a while students with large positive residuals are in the top quintile, having received a . Coefficients in this regression are reported relative to the third quuintile, who experienced the median shock. The sample is restricted to the treatment group for this regression. All regressions follow the controlled specification described in the notes of previous tables. All test outcomes are standardized to have mean zero and standard deviation one among all HISD fifth graders. *** = significant at 1 percent level, ** = significant at 5 percent level, and * = significant at 10 percent level.

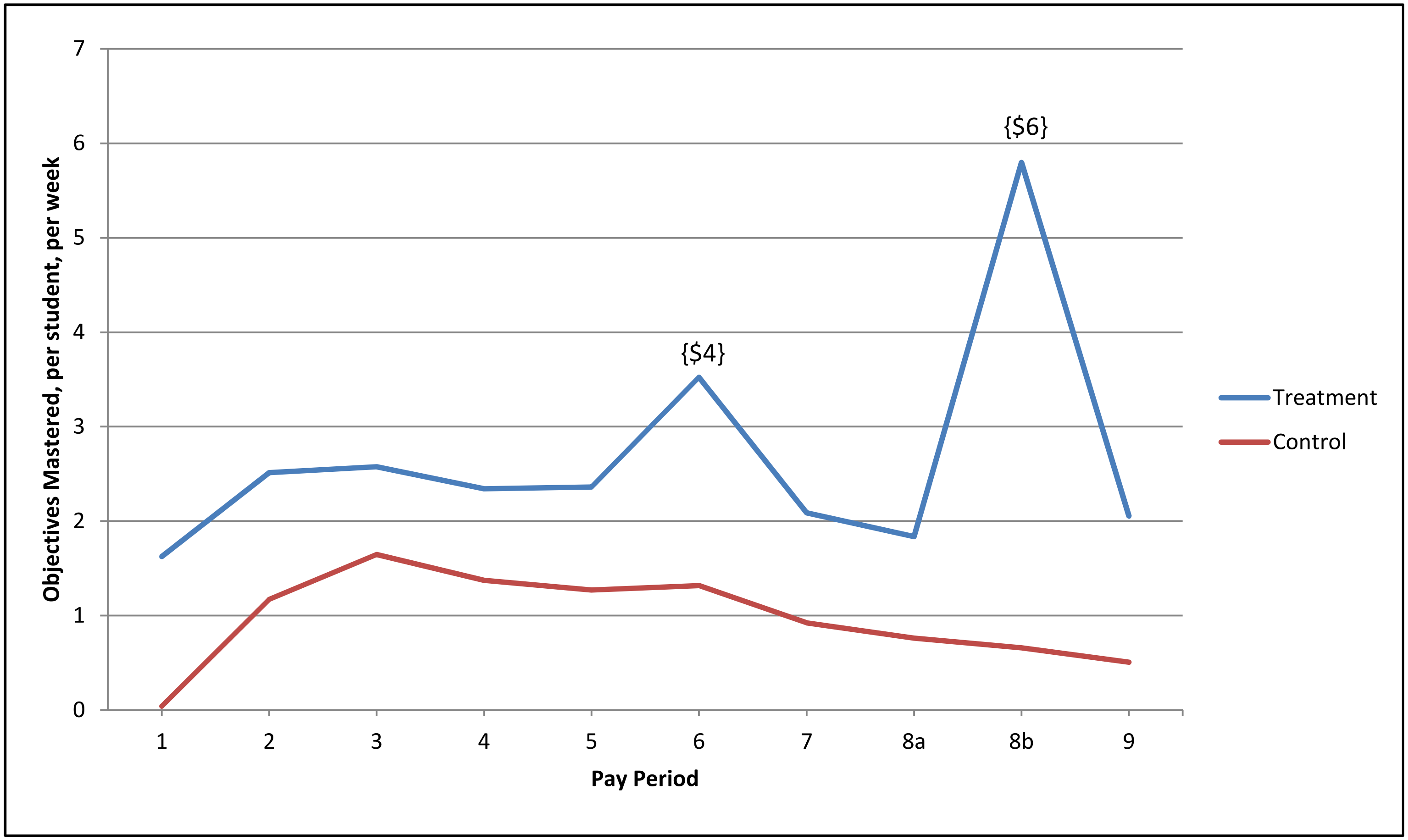

Figure 1: Objectives Mastered by Pay Period

Notes: The vertical axis represents the average number of Accelerated Math (AM) objectives mastered by the average student per week. The horizontal axis represents each pay period. Prices in braces above individual points indicate changes in the price paid to treatment students per objective mastered in AM. If no price is indicated in braces above a point, treatment students were paid $2 per objective during that pay period. Control students were never paid at any price level.

Appendix Table 1 - Mean Effect Sizes (ITT and TOT Estimates)

	First Stage	ITT	TOT
A. Direct Outcomes			
Objectives Mastered	0.954***	1.087***	1.139***
	(0.003)	(0.031)	(0.032)
	3292	3292	3292
Conferences Attended	0.956***	1.572***	1.644***
	(0.003)	(0.099)	(0.103)
	2052	2052	2052
B. Indirect Outcomes			
State Math 10-11	0.957***	0.081***	0.085***
	(0.002)	(0.025)	(0.026)
	3128	3128	3128
State ELA 10-11	0.956***	-0.077***	-0.080***
	(0.003)	(0.027)	(0.028)
	3108	3108	3108
Aligned State Math 10-11	0.959***	0.137***	0.142***
	(0.002)	(0.028)	(0.030)
	3090	3090	3090
Unaligned State Math 10-11	0.959***	0.026	0.027
	(0.002)	(0.030)	(0.032)
	3090	3090	3090
Parents check HW more	0.957***	0.071***	0.074***
	(0.003)	(0.027)	(0.028)
	2041	2041	2041
Student prefers Math to Reading	0.959***	0.112***	0.117***
	(0.004)	(0.023)	(0.024)
	2356	2356	2356
Parent asks about Math more than Rdg.	0.959***	0.122***	0.127***
	(0.003)	(0.028)	(0.029)
	1908	1908	1908
Attendance 10-11	0.962***	0.050*	0.052*
	(0.002)	(0.027)	(0.028)
	3187	3187	3187
Intrinsic Motivation	0.958***	0.006	0.006
	(0.005)	(0.060)	(0.062)
	2004	2004	2004

Notes: This table reports ITT and TOT estimates of the effects of our aligned incentives experiment on a variety of outcomes. First-stage estimates report the causal effect of the experiment on the percentage of the school year each student spends in a treatment school(number of days present divided by 180), controlling for our full set of covariates. ITT estimates mirror those presented in earlier tables. Treatment-on-Treated estimates use randomized assignment to a treatment school to instrument for time spent in a treatment schools; the estimates can be interpreted as the effect of spending a full year in the treatment school for treated individuals. Standard errors are robust to heteroskedasticity. The construction of each dependent variables is described in the notes of previous tables. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A2: Mean Effect Sizes (Intent to Treat) By Subsample

	Whole Sample	*Gender* Male	Female	p-val	*Race* Black	Hispanic	p-val	*Free Lunch* Yes	No	p-val	*Math Quintile* Bottom	Top	p-val
A. Student Achievement													
State Math 10-11	0.081***	0.106***	0.040		-0.002	0.104***		0.144***	-0.006		-0.004	0.228***	
	(0.025)	(0.035)	(0.037)	0.183	(0.056)	(0.033)	0.101	(0.034)	(0.037)	0.003	(0.049)	(0.082)	0.011
	3128	1636	1491		828	2165		1687	1421		663	428	
State ELA 10-11	-0.077***	-0.067*	-0.090**		-0.069	-0.076**		-0.033	-0.122***		-0.165***	0.023	
	(0.027)	(0.037)	(0.039)	0.678	(0.071)	(0.033)	0.926	(0.038)	(0.041)	0.106	(0.063)	(0.083)	0.060
	3108	1616	1491		821	2151		1677	1411		659	427	
Aligned State Math 10-11	0.137***	0.181***	0.084**		0.021	0.177***		0.186***	0.062		-0.010	0.144***	
	(0.028)	(0.041)	(0.040)	0.086	(0.075)	(0.036)	0.056	(0.038)	(0.044)	0.030	(0.090)	(0.049)	0.118
	3090	1619	1470		808	2148		1661	1409		648	427	
Unaligned State Math 10-11	0.026	0.030	0.007		-0.022	0.043		0.080*	-0.045		-0.032	0.129**	
	(0.030)	(0.042)	(0.045)	0.695	(0.081)	(0.038)	0.460	(0.041)	(0.046)	0.040	(0.086)	(0.055)	0.102
	3090	1619	1470		808	2148		1661	1409		648	427	
B. Survey Outcomes													
Parents check HW more	0.071***	0.066	0.075**		0.014	0.092***		0.042	0.121***		0.006	0.184**	
	(0.027)	(0.041)	(0.037)	0.859	(0.067)	(0.034)	0.282	(0.037)	(0.041)	0.141	(0.069)	(0.085)	0.078
	2041	1008	1030		527	1414		1117	911		387	271	
Prefer Math to Reading	0.112***	0.104***	0.130***		0.065	0.119***		0.154***	0.056*		0.153**	0.082	
	(0.023)	(0.032)	(0.033)	0.565	(0.069)	(0.029)	0.465	(0.033)	(0.033)	0.032	(0.062)	(0.063)	0.396
	2356	1214	1136		575	1656		1252	1087		506	299	
Parents ask more about Math	0.122***	0.088**	0.137***		0.173**	0.118***		0.104***	0.136***		0.032	0.214***	
	(0.028)	(0.041)	(0.038)	0.366	(0.073)	(0.036)	0.488	(0.037)	(0.042)	0.561	(0.068)	(0.077)	0.057
	1908	945	960		480	1334		1052	843		356	259	
Conferences Attended	1.572***	1.691***	1.416***		1.708***	1.474***		1.608***	1.592***		1.492***	1.880***	
	(0.099)	(0.148)	(0.136)	0.159	(0.248)	(0.130)	0.386	(0.132)	(0.155)	0.936	(0.234)	(0.305)	0.271
	2052	1018	1030		526	1424		1127	911		394	270	
Objectives Mastered	1.087***	1.012***	1.159***		0.816***	1.114***		1.096***	1.055***		0.686***	1.660***	
	(0.031)	(0.045)	(0.043)	0.017	(0.045)	(0.045)	0.000	(0.043)	(0.047)	0.519	(0.047)	(0.117)	0.000
	3292	1728	1554		857	2283		1774	1492		694	423	
C. Attendance and Motivation													
Attendance 2010-2011	0.050*	0.056	0.038		0.132**	0.004		0.051	0.031		0.048	0.073	
	(0.027)	(0.039)	(0.037)	0.736	(0.066)	(0.031)	0.073	(0.036)	(0.041)	0.700	(0.066)	(0.066)	0.776
	3187	1658	1528		821	2233		1707	1466		700	428	
Intrinsic Motivation	0.006	-0.061	0.048		-0.134	0.016		-0.057	0.079		0.122	0.189	
	(0.060)	(0.089)	(0.083)	0.359	(0.147)	(0.081)	0.353	(0.087)	(0.088)	0.259	(0.156)	(0.202)	0.772
	2004	1029	969		476	1426		1072	916		400	260	

Notes: This table reports ITT estimates of the effects of the experiment on incentivized and non-incentivized outcomes for a variety of subsamples. All regressions follow the controlled specification described in the notes of previous tables. All test outcomes are standardized to have mean zero and standard deviation one among all HISD fifth graders. *** = significant at 1 percent level, ** = significant at 5 percent level, and * = significant at 10 percent level.

Appendix Figure 1:
Accelerated Math Progress Monitoring Dashboard

Accelerated Math

Practice TOPS Report
for

Printed Friday, February 6, 2009 10:45:20 AM

School: East Elementary School
Class: Math 5A

Teacher: Ms. B. Beck
Grade: 5

Number Correct: 11 / 14 (79%)

Incorrect Responses (3)

Objective		Problem	Your Answer	Correct Answer
158.	Answer a question using information from a Venn diagram	4	B	A
159.	Determine the mode from a graph	7	-	D
159.	Determine the mode from a graph	8	?	C

Objectives on this Practice (4)

Objective		Results		Overall	
158.	Answer a question using information from a Venn diagram	5 / 6	83%	9 / 12	75%
159.	Determine the mode from a graph	4 / 6	67%	6 / 12	50%
112. [c]	Determine an appropriate unit of measure	1 / 1	100%	4 / 4	100%
113. [c]	Convert between customary units of length using fractional amounts	1 / 1	100%	4 / 4	100%

Overall Progress

Average Percent Correct	Marking Period (33% Complete)	School Year (58% Complete)
Practice %:	80 [a]	84 [a]
Test %:	88 [b]	89 [b]
Review %:	82	88

Objective Summary

Ready to Test: 1
Goal for Marking Period: 36
Total Mastered this Marking Period: 12 (33% of Goal)
Total Mastered this Year: 96

Teacher

Parent

Comments:

[a] Includes Exercise Results
[b] Includes Diagnostic Test Results
[c] Review Objectives

Form: 2820
Problems: 1-14
Printed: 2/6/09 10:45:20 AM
Completed: 2/6/09 10:45:08 AM

Appendix Figure 2:
Distribution of Treatment and Control Schools Across Houston

Notes: The background color indicates the poverty rate for each census tract, with darker shades denoting higher concentrations of poverty. T's and C's mark treatment and control schools, respectively.