NBER WORKING PAPER SERIES

THE BEHAVIORALIST GOES TO SCHOOL: LEVERAGING BEHAVIORAL ECONOMICS TO IMPROVE EDUCATIONAL PERFORMANCE

Steven D. Levitt
John A. List
Susanne Neckermann
Sally Sadoff

Working Paper 18165
http://www.nber.org/papers/w18165

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2012

We gratefully acknowledge the leadership and support of our Bloom Township, Chicago Heights and Chicago Public School District partners. We also thank Bruno Frey, Marcin Siwicki and Esjay Yoo for their valuable contributions to the experimental design. Alec Brandon, Patrick Fitz, Trevor Gallen, Sean Golden, David Herberich, Wooju Lee, Ian Muir, Joseph Seidel, Phuong Ta and Timo Vogelsang provided truly outstanding research assistance. The project was made possible by the generous financial support of the Children First Fund, the Kenneth and Anne Griffin Foundation, the Rauner Family Foundation, and the Spencer Foundation. This research has been conducted with IRB approval. Please direct correspondence to Sally Sadoff. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2012 by Steven D. Levitt, John A. List, Susanne Neckermann, and Sally Sadoff. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance

Steven D. Levitt, John A. List, Susanne Neckermann, and Sally Sadoff

NBER Working Paper No. 18165
June 2012
JEL No. C9,C93,H75,I20

ABSTRACT

A long line of research on behavioral economics has established the importance of factors that are typically absent from the standard economic framework: reference dependent preferences, hyperbolic preferences, and the value placed on non-financial rewards. To date, these insights have had little impact on the way the educational system operates. Through a series of field experiments involving thousands of primary and secondary school students, we demonstrate the power of behavioral economics to influence educational performance. Several insights emerge. First, we find that incentives framed as losses have more robust effects than comparable incentives framed as gains. Second, we find that non-financial incentives are considerably more cost-effective than financial incentives for younger students, but were not effective with older students. Finally, and perhaps most importantly, consistent with hyperbolic discounting, all motivating power of the incentives vanishes when rewards are handed out with a delay. Since the rewards to educational investment virtually always come with a delay, our results suggest that the current set of incentives may lead to underinvestment. For policymakers, our findings imply that in the absence of immediate incentives, many students put forth low effort on standardized tests, which may create biases in measures of student ability, teacher value added, school quality, and achievement gaps.

Steven D. Levitt
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
slevitt@midway.uchicago.edu

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Susanne Neckermann
ZEW (Centre for European Economic Research)
neckermann@zew.de

Sally Sadoff
Rady School of Management
University of California San Diego
sadoff@uchicago.edu

1 Introduction

Behavioral economics has now gone beyond mere academic curiosity, touching nearly every field in economics. Theorists are recognizing behavioral regularities that lie outside of the standard paradigm in their models, empiricists are taking new behavioral predictions to the lab and field, and policymakers are increasingly recognizing the power of psychology when crafting new legislation. One area where behavioral economics has made only limited inroads, however, is in education circles. This is puzzling since it is an area where the insights gained from behavioral economics might be especially great.

In this study, we use a series of field experiments to explore how behavioral economics can be leveraged to improve student performance. Our experiments revolve around three major behavioral tenets. First, some people have reference-dependent preferences, wherein utility is determined not just by absolute levels of consumption, but also by consumption relative to a reference point. For instance, in certain cases, such people will exhibit behavior consistent with a notion of *loss aversion*, an insight gained from Kahneman and Tversky's (1979) prospect theory.

Second, non-material rewards, for example, in the form of awards and trophies can have considerable motivational power (Kosfeld and Neckermann 2011). Such rewards derive their motivatational power from a variety of mechanisms including status, self-image concerns, and relative performance feedback that have been shown to affect behavior.[1]

Finally, some people have hyperbolic preferences, overweighting the present so much that future rewards are largely ignored (see, e.g., Strotz 1955, Laibson 1997). Such preferences can lead to underinvestment when (as in education) the returns to achievement are largely delayed.[2]

[1]See Ball et al. (2001) and Huberman et al. (2004) and on status; Blanes i Vidal and Nossol (2011), Tran and Zeckhauser (2009) and Barankay (2011) on relative performance feedback; and Ariely et al. (2009) and DellaVigna et al. (2012) on image motivation and social pressure.

[2]Previous studies find a negative correlation between hyperbolic discount rates and educational outcomes (Kirby et al. 2002, Kirby et al. 2005). Similarly, Mischel et al. (1989) find that measures of ability to delay gratification in early childhood are predictive of longer-term academic achievement.

We investigate each of these three areas of behavioral economics using field experiments conducted over multiple sites and years. This permits a glimpse of behavior not only within the experimental period but for months afterwards. Our field experiments include over 7,000 elementary and high school students in three school districts in and around Chicago. The typical study reports findings from a single experiment without any replications to examine transferability to different settings and scales. This paper addresses both questions by studying the impact of various incentive designs in several settings, among a wide age range of students and in school districts of very different size.[3]

In our baseline setup, students are offered cash or non-pecuniary rewards for an improvement in test scores. The tests last between 15 - 60 minutes, yielding a high hourly wage (particularly in the highest financial incentive group) that is likely quite salient among our subject pool of low-income children and adolescents. We investigate the effectiveness of low and high financial incentives ($10, $20) and compare these to the impact of non-monetary rewards in the form of a trophy for achievement. These incentives are presented in either the gain or the loss domain. In addition, we test directly the importance of discount rates by offering incentives either immediately after the test or with a delay (a month after the test).

Importantly, the incentives are announced immediately before the test (with no advance notice). This design allows us to isolate the effect of incentives on performance solely through inducement of greater short-run effort – avoiding confounding due to discount rates or human capital accumulation (i.e., studying for the test).

We find that incentives affect student performance, although there is substantial variation in the effectiveness of rewards across settings. Yet, one robust result is that incentives framed as losses have consistently large effects relative to other educational interventions ($0.08 - 0.17$ standard deviations). These effect sizes are comparable to those achieved through a one-

[3]In a similar vein, Braun et al. (2011) test a single performance pay incentive among 2,600 students in 59 schools and seven states. Fryer (2011) reports on a series of financial incentive programs carried out in a number of large American school districts (but does not compare different incentive designs within a single setting).

standard deviation increase in teacher quality (e.g., Rockoff 2004, Rivkin et al. 2005) or a one-third reduction in class size (Krueger 1999). We find mixed evidence on the impact of incentives framed as gains with large effects in two school districts ($0.12 - 0.46$ standard deviations) and no effects in the third. We also find that that while older students are more responsive to financial incentives, non-financial incentives are as effective as financial incentives among younger students (and thus more cost-effective). Finally, we find that non-immediate incentives (rewarded a month after the test) have no effect.

The design also allows us to uncover some of the underlying heterogeneities that drive the overall effectiveness of reward schemes: younger children are more responsive than older children, for whom only the high financial incentive presented as a loss affected performance. Moreover, boys are more responsive to short-term incentives than girls. Overall, effects are more pronounced for math tests than for reading tests.

Our results suggest that in the absence of immediate incentives, many students put forth low effort on the standardized tests that we study. These findings have important implications for policymakers because standardized assessment tests are often high-stakes for teachers and principals (e.g., as determinants of school resources), but low-stakes for the individual students choosing to exert effort on the test. Low baseline effort among certain groups of students can create important biases in measures of student ability, teacher value added, school quality, and achievement gaps.[4] If delays in rewards reduce student effort in this context, it would seem likely that the general pattern of delayed rewards in the educational setting (e.g., increased earnings associated with school attainment accrue only with lags of years or even decades) could induce sub-optimal effort. Contrary to a widespread concern, we also do not find that incentives have a detrimental effect on performance in

[4]Baumert and Demmrich (2001) and Braun et al. (2011) make a similar argument based on their findings and review the literature on achievement gaps due to differential motivation. In a similar vein, Jacob (2005) uncovers evidence that differential effort on the part of students can explain the otherwise puzzling divergence over time in the performance of students in the Chicago Public Schools on high-stakes versus low-stakes tests. It appears that CPS teachers and administrators became increasingly successful over a period of years at convincing students to take the high-stakes test seriously, but that same effort did not spill over to the low stakes state-administered tests.

subsequent tests.

The remainder of the paper is organized as follows. Section II summarizes the underlying theoretical framework that motivates our design. Section III describes the experimental design and implementation. Section IV discusses the main results and potential sources of heterogeneity. Section V concludes with a discussion of the broader implications of the findings.

2 Theoretical Framework

Although financial incentives are ubiquitous in modern society, direct financial incentives have not traditionally been used to motivate student effort and performance. In recent years, however, monetary rewards have begun to attract attention from educators and policymakers. Recent programs have conditioned monetary rewards on a variety of measures including school enrollment, attendance, behavior, grades, test performance, and matriculation.[5]

Although results have varied across settings, financial incentives have generally been associated with modest positive improvements in student performance. Typically, these incentive programs have remained in place for an extended period of time (e.g. a school year) with the goal of affecting student behavior over the duration of that time horizon. Numerous studies however find that children and adolescents tend to exhibit high discount rates and have difficulty planning for the future (e.g., Gruber 2001, Bettinger and Slonim 2007, Steinberg et al. 2009). They may therefore respond more strongly to rewards with very short time horizons compared to incentives extending over several months or years.

At the same time, behavioral anomalies, such as the endowment effect (Thaler 1980), status quo bias (Samuelson and Zeckhauser 1988), and observed divergences of willingness to pay and willingness to accept measures of value (Hanemann 1991) have arisen in the broader

[5]Examples include Progresa in Mexico which offered incentives for school enrollment and attendance (Schultz 2004, Behrman et al. 2005). A similar conditional cash transfer program was instituted in Colombia (Barrera-Osorio et al. 2008). Other programs have based rewards on overall school performance (see Angrist et al. 2006, Levitt et al. 2010, Leuven et al. 2010, Fryer 2011).

literature. These examples of reference-dependent decision making are broadly consistent with a notion of *loss aversion*, an insight gained from Kahneman and Tversky's (1979) prospect theory. A more recent branch of behavioral economics has explored the effectiveness of non-financial rewards (e.g., Frey 2007, Bradler et al. 2012, Ashraf et al. 2012).

To formalize these ideas, we consider a representative agent who derives benefits and costs as follows in period t :

$$v_t = v(e_t, r_t, r_t^r) = \pi(e_t)[u(r_t) + R(r_t, r_t^r)] + [1 - \pi(e_t)]R(0, r_t^r) - c(e_t)$$

where individuals receive rewards r with probability π (and receive 0 with probability $1-\pi$), u is utility over rewards, R is the value function of prospect theory and c is the cost of effort e. Let $\pi(.)$ be increasing and concave in e, $u(.)$ be increasing and concave in r, $c(.)$ be increasing and convex in e, and we normalize $u(0) = 0$. We define utility derived in relation to a reference point r^r, $R(.)$:

$$R(r, r^r) = \begin{cases} g(r - r^r), & if \quad r \geq r^r \\ h(r - r^r), & if \quad r < r^r \end{cases}$$

where g is increasing and concave, h is increasing and convex and we normalize $g(0) = 0$. Rewards r are a weighted sum of the material benefits w and non-pecuniary benefits b that the agent receives

$$r = \mu w + \eta b$$

where μ and η represent the weights of material and non-pecuniary benefits for the individual's utility, respectively. As aforementioned, examples of non-pecuniary benefits include status, a positive self-image, enjoyment of the task, and the trophy value of certain rewards.[6]

A person in period t cares not only about her present instantaneous utility v_t but also

[6]For individuals that care about status and a positive self-image, non-pecuniary gifts carry additional utility when they remind oneself and others of a special achievement of the individual (see, e.g., Loewenstein and Issacharoff (1994) on the trophy value of rewards and Bénabou and Tirole (2006) on self-signaling).

about her future instantaneous utilities.[7] $V^t(v_t, v_{t-1}, \dots v_T)$ represents a person's intertemporal preferences from the perspective of period t, where V^t is continuous and increasing in all its components and has the following form

$$V^t(v_t, v_{t+1}, \dots v_T) \equiv \delta^t v_t + \beta \sum_{r=t+1}^{T} \delta^t v_t \quad \text{for all } t, \text{ where } \quad 0 < \beta, \delta \leq 1.$$

Individuals exhibit hyperbolic discounting and are present-biased if $\beta < 1$ (δ represents the constant discount rate). If students have such preferences, then their utility from immediate rewards is far greater than the utility they receive from the same reward in the future.

In period $t = 0$, a student chooses effort to maximize V^0. If rewards are immediate, then the student's objective function is

$$\max_{e_0} \ \pi(e_0)[u(r_0) + R(r_0, r_0^r)] + [1 - \pi(e_0)]R(0, r_0^r) - c(e_0)$$

If rewards are delayed by one period, then the student's objective function becomes

$$\max_{e_0} \ \beta\delta \left\{\pi(e_0)[u(r_1) + R(r_1, r_1^r)] + [1 - \pi(e_0)]R(0, r_1^r)\right\} - c(e_0)$$

where $r_0 = r_1 \equiv r$ and $r_0^r = r_1^r \equiv r^r$. The respective first order conditions for these equations are $c' = \pi'(u^r + R^r - R^0)$ and $c' = \beta\delta[\pi'(u^r + R^r - R^0)]$.[8] The cost of effort is equivalent in both cases because effort is always exerted immediately. However, the benefit of the delayed reward is discounted by $\beta\delta \leq 1$, thus reducing optimal effort under delayed rewards relative to optimal effort under immediate rewards.

We can similarly compare the objective function for immediate rewards presented as gains[9]

$$\max_{e} \ \pi(e)[u(r) + R(r, 0)] + [1 - \pi(e)]R(0, 0) - c(e)$$

[7] We follow the framework developed in O'Donoghue and Rabin (1999).

[8] $u^x \equiv u(x)$ and $R^x \equiv (x, r^r)$

[9] Hereafter t=0 unless otherwise noted.

to the objective function for immediate rewards presented as losses

$$\max_e \ \pi(e)[u(r) + R(r, r^r)] + [1 - \pi(e)]R(0, r^r) - c(e)$$

where $r = r^r$. The respective first order conditions for these equations are $c' = \pi'(u^r + g(r))$ and $c' = \pi'(u^r - h(-r))$. If losses are felt more strongly than gains $-h(-r) > g(r)$, then optimal effort will increase if rewards are framed as losses rather than gains. Estimates of the ratio of h and g (when linearity is assumed) have found $-h(-x)/g(x) \cong 2$ (see Tversky and Kahneman 1991). In the spirit of this finding, if students are loss-averse, then the negative utility a student receives from a loss of x is greater in magnitude than the positive utility she receives from a gain of x for any positive x. And thus optimal effort will increase when rewards are framed as losses rather than gains.

Finally, we compare the objective function for immediate monetary rewards r^m

$$\max_e \ \pi(e)[u(r^m) + R(r^m, r^r)] + [1 - \pi(e)]R(0, r^r) - c(e)$$

to the objective function for immediate non-pecuniary rewards r^n

$$\max_e \ \pi(e)[u(r^n) + R(r^n, r^r)] + [1 - \pi(e)]R(0, r^r) - c(e)$$

where r^n and r^m are of equivalent cash value. The respective first order conditions for these equations are $c' = \pi'[u(r^m) + R(r^m, r^r) - R^0]$ and $c' = \pi'[u(r^n) + R(r^n, r^r) - R^0]$. Optimal effort will be higher under non-pecuniary rewards than under monetary rewards when $r^n = \mu w^n + \eta b$ exceeds $r^m = \mu w$, where w is the cash value of the rewards and $w^n \leq w$ is the material benefit of the non-pecuniary prize.[10] This occurs when $\mu(w - w^n) < \eta b$.

Accordingly, depending on the relative weight that students place on non-pecuniary ben-

[10]We assume that cash prizes provide zero non-pecuniary benefits $b^m = 0$. The material benefit of non-pecuniary rewards is never greater than and typically less than their cash value $w^n \leq w$ because they are less fungible than cash (Waldfogel 1993).

efits ($\eta \gg 0$ represents a strong preference for non-pecuniary rewards), the utility derived from gifts or trophies potentially exceeds the value of monetary rewards of equivalent cost. The utility weights μ and η may also vary across individuals. For example the utility weighting of non-pecuniary benefits η may decline with age, in which case non-pecuniary rewards are relatively more effective among younger students.

In all treatments, we announce the rewards immediately before the incentivized test (with no advance notice). This ensures that as in the equations above, students are choosing one-time effort in the immediate period only – i.e., there is no delay between effort exertion and immediate rewards. This feature allows us to ignore time discounting of effort as well as human capital gains that may accrue from effort in previous periods (e.g., studying for the test), and allows us to focus on a decision that is most in concert with the theory tested.

Most previous programs that have rewarded incentives based on test performance have announced the incentive well in advance of the test using high school exit and achievement exams in Israel (Angrist and Lavy 2009) and Texas (Jackson 2010); and standardized tests for elementary/middle school students in Kenya (Kremer et al. 2009), India (Berry 2011), Ohio (Bettinger 2010) and New York City (Fryer 2011). Studies that have announced incentives immediately before the test have typically distributed rewards with a delay. The evidence on such delayed rewards is mixed. O'Neil et al. (1997, 2004) find that delayed financial incentives can increase eighth grade test scores but have no effect on twelfth grade test scores, even at very high levels (up to $100 on a 10 question test).[11] In a similar design, Baumert and Demmrich (2001) find no effects of financial incentives on ninth grade test scores. These studies also find no treatment effects from non-financial incentives including feedback, ranking, goal setting, achievement certificates, and test scores counting towards classroom grades.[12]

[11] O'Neil et al. (2004) also offered an immediate incentive of up to $20 based on answering 2 practice questions correctly. However, all participants in both the incentive and control groups answered the questions correctly, precluding any measurement of a treatment effect.

[12] To our best knowledge, a study produced concurrently to ours - Braun et al. (2011) - is the only other study to announce the incentive immediately before the test and distribute the reward immediately after the test. They offer a performance-based incentive of up to $30 to eighth and twelfth graders on a low stakes

3 Experimental Design and Implementation

The field experiment was carried out in six waves in three low-performing school districts in and around Chicago: Bloom Township (Bloom), Chicago Heights (CH) and Chicago Public Schools (CPS). The first two waves were conducted in winter and spring 2009 among high school sophomores at one high school in Bloom. The third wave took place in spring 2010 with a new cohort of Bloom sophomores. The fourth wave also took place in spring 2010 among 3rd-8th graders in seven elementary schools in Chicago Heights. The final waves scaled up the Bloom and Chicago Heights experiments and were conducted in 26 CPS elementary schools among 2nd-8th graders in fall 2010 and winter 2011.[13]

The field experiment took place during regularly scheduled sessions of standardized diagnostic tests. These are low-stakes tests that students do not generally prepare for or have any external reason to do well on. Students generally take the tests three times a year in the fall, winter, and spring.[14] They are computer-based and last between 15-60 minutes with students' results available immediately after the test ends.[15]

In each session, immediately before testing began, the test administrator announced the incentive and told students that they would receive the reward immediately (or a month) after the test ended if they improved upon their baseline score from a prior testing session. Immediately after the test ended, we handed out rewards to qualifying students, except in the case of delayed rewards which were distributed a month after testing.[16] Students received no advance notice of the incentives prior to the testing sessions.[17]

standardized test and find positive and significant treatment effects compared to a control group which received no incentive and a "fixed incentive" group which received $20 regardless of performance.

[13]Bloom and Chicago Heights are small school districts with approximately 3,000 students each. In contrast, CPS is the third largest school district in the U.S. with approximately 400,000 students.

[14]In Chicago Heights, students also take a pre-test at the beginning of the year.

[15]In Bloom, the experiment took place during the STAR Reading Assessment, which is adaptive and lasts about 15 minutes. In Chicago Heights, the experiment took place during the math portion of the ThinkLink Predictive Assessment Series, which lasts about 30 minutes. In CPS, the experiment took place during either the math or reading portion of the Scantron Performance Series, which each last about 60 minutes.

[16]In CPS, about one-fifth of classes did not complete testing in a single session due to time constraints. In these cases, we returned to the school after every student had completed the test. Excluding these classes from the analysis does not affect the results.

[17]One week before testing, we sent home a consent form to parents stating that we would like their child

Incentivized students were offered one of the following rewards: financial low ($10 cash), financial high ($20 cash) or non-financial (trophy). In the loss condition (financial high and non-financial) students received the reward at the start of the testing session and were informed that they would keep the reward if they improved (and that they would lose the reward if they did not improve). Students also filled in a sheet confirming receipt of the reward (and in CPS what they planned do with it) and kept the reward at their computer during testing. In the control groups, the test administrator either did not make any announcement (control - no statement) or encouraged students to improve on the test but did not offer any incentive to do so (control - statement).[18] This allows us to test whether there are effects due to the presence of the experimenters (we did not attend "no statement" treatments) or of merely requesting that the student improve. Scripts for the different treatments can be found in Appendix A. An overview of the treatments conducted is presented in Table 1.[19]

We randomized at the level of English class (Bloom) or school-grade (CH and CPS) and blocked the randomization on average baseline score, school (CH and CPS), grade (CH and CPS), and race/ethnicity (CH).[20] In cases where students participated in two waves (Bloom

to participate in a study to be conducted during the upcoming test. And, that their child could receive financial or non-financial (where applicable) compensation for their participation. We did not specify the incentives and we sent the same consent form to the treatment and control groups. In Bloom and Chicago Heights, parents only needed to sign the consent form if they did *not* want their child to participate in the study. Less than 1% of parents opted out by returning the form. In CPS, parents needed to sign the consent form in order for their child to participate. 57% of parents returned the signed consent form prior to the fall session and 71% of forms were returned prior to the winter session. In order to participate, students in all sessions that we attended also signed a student assent form immediately before they took the test. All students opted into the study by signing the assent form. The analysis only includes students who met the consent criteria prior to treatment.

[18]In Chicago Heights, a second financial low (comparison) treatment and a second control-statement (comparison) treatment added a statement that we would compare a student's improvement to three other students with similar past scores. The non-financial treatment added a statement that we would take a photo of qualifying students and post it in their school. In CPS, control - statement students were additionally told (as incentivized students were) that they would learn their scores either immediately or with a one month delay (control - statement - delayed) after testing.

[19]The various waves included additional incentive treatments which are available in an online appendix. The main part of this paper only includes the incentives that are common across the settings.

[20]In Bloom, we blocked on baseline reading score. If the baseline score was not available, we blocked classes by their track: regular, remedial or honors. In CH and CPS, we blocked on baseline math and reading scores.

2009 and CPS 2010/2011), we re-randomized for the second wave.[21] Thus, some students received the same treatment in both sessions, while others received a different treatment in the two sessions. In the two cases where students received incentives in a previous session (Bloom spring 2009 and CPS winter 2011) there was no particular reason for students to expect the experiments to continue, or if the experiments did continue, that they would receive a particular incentive. It is possible, however, that students anticipated there would be incentives in their second testing session. The results presented below are robust to restricting the sample to first-time incentives (i.e., to students in their first testing session and those in their second session who were in the control group in the first session).

Tables 2a-2c report summary statistics by treatment group for pre-treatment characteristics in Bloom (2009 and 2010), Chicago Heights (2010) and CPS (2010 and 2011). The pre-treatment characteristics include baseline score, grade (CH and CPS), test subject (CPS) and the following demographics: gender, race/ethnicity, free/reduced lunch status and (in CH and CPS) eligibility for an Individualized Education Plan (IEP).[22] While the groups are generally balanced, the tables indicate the presence of some significant differences between incentive and pooled control (statement and no statement) groups, with standard errors clustered by class (Bloom) or school-grade (CH and CPS). In Bloom (Table 2a) the only significant differences are the proportion of black and Hispanic students in the financial low ($10) treatment. In Chicago Heights (Table 2b) all three treatment groups have a significantly lower average grade than control. The financial high and non-financial treatment groups have significantly lower proportions of black students and significantly higher proportions of Hispanic students than control. In CPS (Table 2c) the various treatment groups are balanced on average grade and baseline score (the non-financial incentive group has

[21] In the second CPS wave, we additionally blocked on treatment received in the first wave, math and reading scores in the first wave, and treatment received in a separate intervention that took place between the two waves.

[22] Baseline test score is a standardized pre-treatment test score. In Bloom 2009, fall 2008 serves as the baseline. In Bloom 2010, fall 2009 serves as the baseline. In Chicago Heights, winter 2010 serves as the baseline. In CPS, spring 2010 serves as the baseline. Eligibility for free/reduced lunch is a proxy for family income. Individualized Education Plans (IEPs) provide additional services to struggling students. IEP status was not available for Bloom students.

higher baseline scores than control significant at the $p < 0.1$ level). There are statistically significant differences (both positive and negative) in the proportion of math tests, as well as demographic measures in some groups. As shown below, the results are robust to including controls for baseline performance and other pre-treatment characteristics.

4 Results

The following results estimate treatment effects on test score improvement in each of our settings: Bloom (2009 and 2010), Chicago Heights (2010) and CPS (2010 and 2011).[23] The dependent variable in all regressions is test score improvement (in standard deviation units) with standard errors clustered by class (Bloom) or school-grade (CH and CPS).[24] For each setting, we first present treatment effect estimates absent any controls except for the session of the experiment (Bloom and CPS). The second column for each setting adds controls for baseline score in the tested subject (score, score squared and score cubed), past treatment (whether the student was incentivized in a previous wave in Bloom and CPS), test subject (CPS), school and grade (CH and CPS), teacher fixed effects (Bloom), and demographics (gender, race/ethnicity, free/reduced lunch eligibility and (in CH and CPS) IEP status).[25]

The omitted category in every regression is the pooled control (statement and no statement) group for the relevant setting. There are no significant differences in performance between the control subgroups and pooling does not affect the results. This suggests that the treatment effects are due to the incentives rather than the presence of the experimenters or the mere encouragement to improve.

[23] An analysis of the individual waves (i.e., without pooling) yields similar results.

[24] Improvement is measured as the difference between the standardized outcome score and the standardized score students were told to improve on (baseline score in Bloom and CH, and prior session score in CPS). Scores are standardized to have mean zero and standard deviation equal to 1. In Bloom, we standardize scores within each testing period using the full sample of Bloom students. In Chicago Heights, we standardize scores within each grade and testing period using the full sample of Illinois students. In CPS, we standardize scores within each grade, subject and testing period using the full population of CPS students.

[25] The results below are robust to bootstrapping clustered standard errors in regressions with fewer than forty-two clusters.

Result 1: Monetary incentives matter

We first examine whether financial incentives can improve performance by inducing greater student effort. Table 3 reports the estimated effects of immediate incentives on test score improvement by setting: Bloom (columns 1 and 2), Chicago Heights (columns 3 and 4) and CPS (columns 5 and 6). We find that sufficiently high financial incentives have a substantial effect. The point estimates of the $20 incentives (framed either as gains or losses) are consistently positive and generally statistically significant at conventional levels, ranging from $0.057 - 0.1$ standard deviations in CPS to $0.37 - 0.46$ standard deviations in Chicago Heights. The magnitude of the impact is economically significant as well: the middle range of effects in Bloom ($0.12 - 0.2$ standard deviations) are equivalent to about 5-6 months' worth of learning on the test.[26] The large effects of these relatively modest financial incentives suggest that at baseline this population of students puts forth low effort in response to low (perceived) returns to achievement on standardized tests.

We also find that the size of the financial incentive matters. Within each setting, the point estimates for the $10 incentives are lower than the point estimates for the $20 incentives, with these differences being significant at the $p < .05$ level in Bloom and CPS. Overall, the low financial incentives yields mixed results with large and significant effects in Chicago Heights ($0.21 - 0.24$ standard deviations) and no effect in Bloom and CPS where point estimates are negative. As far as we know, ours is the first study to demonstrate that student responsiveness to incentives is sensitive to the size of the reward. These findings suggest that for some students effort costs may be relatively high – in Bloom for example, students were willing to exert significant additional effort in response to an hourly wage of approximately $80 but not in response to a $40 per hour wage.[27]

[26]The month equivalent measure is based on the STAR Reading Assessment Instructional Reading Level. The Instructional Reading Level is the grade level at which a student is at least 80% proficient. An IRL score of 6.6 (the average fall baseline score) indicates that a student is reading at the equivalent of 6th grade and 6 months (with 9 months in a school year).

[27]As we discuss below, it may also be the case that relatively low financial incentives crowd out intrinsic motivation yielding smaller net effects.

Result 2: Incentives are more powerful when framed as losses

We next examine whether insights from behavioral economics can improve the effectiveness of incentives. The first result that stands out is that the effects of incentives framed as losses are more robust than the effects of equivalent incentives framed as gains. The loss treatments (either financial or non-financial) have consistently significant effects: $0.15 - 0.17$ standard deviations in Bloom and $0.08 - 0.12$ standard deviations in CPS. We find mixed evidence on the impact of incentives framed as gains, with large effects in Bloom ($0.12 - 0.20$ standard deviations) and Chicago Heights ($0.26 - 0.46$ standard deviations), but no effects in CPS.[28] These results suggest that incentives are more powerful when framed as losses, particularly in settings where traditional rewards have little or no impact.[29] While numerous laboratory studies have demonstrated such effects, ours is among the first to provide evidence for loss aversion in the field.[30]

Result 3: Non-financial incentives impact performance, but are more potent amongst younger students

Turning to our second behavioral intervention, we compare the effects of non-pecuniary rewards to both low and high monetary rewards, which allows us to price out the effects of non-financial incentives. Within each setting, the point estimates for non-pecuniary rewards (framed as gains or as losses) generally fall between those for the $10 and $20 treatments and are nowhere significantly different from the effects of the high financial incentive. Framed as gains, non-financial rewards have large effects in Chicago Heights ($0.26 - 0.32$ standard deviations) and no effect in CPS. However, framed as losses non-financial incentives have

[28] In CPS, the effect of incentives framed as gains is significant at the $p < 0.05$ level once controls are included.

[29] In addition to framing and loss aversion, the loss treatments may also make the reward more salient and increase students' trust and subjective beliefs with respect to the actual payout of these unusual incentives.

[30] As far as we know, Hossain and List (2009) is the only previous study to test loss aversion in the field, finding that framing bonuses as losses improves the productivity of teams in a Chinese factory. In studies run concurrently to ours, Fryer et al. (2012) find that framing bonuses as losses improves teacher performance while List and Savikhin (2012) find no framing effects for student incentives to make healthy food choices. Krawczyk (2011) tests the effect of framing on risk taking on a final exam and finds no effect – the study does not examine the effect of framing on effort or overall performance.

significant effects ($0.08 - 0.12$ standard deviations) that are similar in size to those of the financial loss treatment and are significantly different from the \$10 incentive at the $p < 0.05$ level. Typically, the material cost of non-financial incentives is low – in our case, one trophy cost approximately \$3. Hence, non-financial incentives are a potentially much more cost effective way of improving student performance than is paying cash, particularly in combination with the loss frame.

We introduced non-financial incentives in the elementary context under the expectation that younger children may be relatively more responsive to non-financial rewards than older students, as they are less familiar with cash and might be more sensitive to framing effects of non-pecuniary rewards. Table 4 estimates treatment effects separately for younger students (grades 2-5) and older students (grades 6-8) in CPS.[31] Overall, younger students are more responsive to incentives with large positive and significant effects in all treatments when controls are included (except financial low). Non-financial rewards framed as a loss work best, increasing performance by $0.18 - 0.25$ standard deviations. Older students, in comparison, only respond to financial incentives framed as a loss, which increase student performance by $0.12 - 0.13$ standard deviations. Within each age group, the effects of financial incentives and non-financial incentives (both framed as losses) are significantly different at $p < 0.05$ level. Hence, non-financial incentives may be a cost-effective alternative to monetary rewards but only with younger children. Non-pecuniary incentives are also attractive because schools tend to be more comfortable rewarding students with trophies, certificates and prizes than they are with using cash incentives.

Result 4: Rewards provided with a delay have little impact on student performance

The effects of the immediate rewards – approximately $0.1 - 0.2$ standard deviations – are larger than those typically found in previous studies of student incentives. As we discussed above, this divergence may be the result of students highly discounting rewards that are

[31] The sample size in Chicago Heights does not allow us to separately estimate treatment effects by age group.

received with even a short delay from the time they must exert effort. In order to explore the importance of discount rates, we implemented a delayed version of the four effective treatments (financial high, non-financial, financial loss and non-financial loss). In delayed treatments, students were informed that they would receive the reward one month after the test (rather than immediately after the test). Table 5 contrasts our findings for immediate rewards with those from rewards that are handed out with a delay. Columns 1 and 2 replicate the estimates for immediate treatment effects from Table 3 (excluding the $10 incentive). Columns 3 and 4 estimate the effects of delayed rewards.

None of the estimated effects of the delayed rewards are statistically significant and more often than not are negative in sign. In this context, students behave as if they have very high discount rates. While this is in line with previous research highlighting the high discount rates of children, it poses a challenge for educators and policymakers. Typically, the results of the state-wide assessments are only available 1-2 months after the administration of the tests. More broadly, if similar discount rates carry over to other parts of the education production function, our results suggest that the current set of incentives may be leading to underinvestment in human capital.

Result 5: Treatment effects vary by age, test subject and gender

As noted above, younger students are more responsive to incentives than older students. Tables 6 and 7 explore heterogeneity in response to incentives along two further dimensions: test subject and gender.[32]

Table 6 presents treatment effects separately for reading and math in CPS. In line with previous evidence, we find that incentives have larger effects on math than reading (e.g., Bettinger 2010, Fryer et al. 2012).[33] The estimated effects in math are all large and positive with the $20 and loss treatments statistically significant. The point estimates

[32]We also examine heterogeneous effects by ability (above or below median performance on baseline test) as well as race, and find no systematic differences. Results are available upon request.

[33]In general, education interventions often have greater impact on math achievement than on reading (e.g., Decker et al. 2004, Rockoff 2004, Jacob 2005, Dobbie and Fryer 2011).

in math ($0.08 - 0.35$ standard deviations) approach those we found in Chicago Heights elementary schools where only math was incentivized. In reading, on the other hand, only non-financial loss has a positive impact on student performance while financial low even decreases performance.

Table 7 examines treatment effects by gender in Bloom and CPS.[34] In both settings, treatment effects are larger for boys than for girls. The estimated effects for boys are all positive and significant (except financial low) with no consistently significant effects among girls. This is in line with the literature showing that boys are more responsive to short-term incentives than girls, which may be due in part to gender differences in time preferences.[35] These results also suggest that girls may be more intrinsically motivated than boys and thus more sensitive to crowding out, which we discuss in more detail below.[36]

Result 6: The introduction of rewards does not crowd-out future effort

The use of financial incentives in the education context has been sharply criticized. Theoretically, the most compelling of these criticisms is that extrinsic rewards crowd out intrinsic motivation, rendering such approaches ineffective in the short run, and potentially detrimental in the long run if intrinsic motivation remains low after the monetary incentives have been removed.[37] However, on tasks where intrinsic motivation is already low or zero, external rewards are less likely to have such negative long-term effects.[38] It is also worth noting that

[34]Again, the sample size in Chicago Heights is too small to conduct a similar analysis.

[35]Evidence on the effect of incentives by gender is mixed with longer term studies tending to find larger effects on girls (e.g. Angrist et al. 2009, Angrist and Lavy 2009) and shorter term studies finding larger effects among boys, particularly in the context of competition (Gneezy and Rustichini 2003, 2004). Bettinger and Slonim (2007) and Castillo et al. (2011) find that boys are more impatient than girls.

[36]There is mixed evidence on gender differences in intrinsic motivation. Baumert and Demmrich (2001) as well as Vallerand et al. (1992) report that girls are more instrinsically motivated than boys. However, Vallerand et al. (1994) do not find a gender difference. Skaalvik and Skaalvik (2005) as well as Wolters and Pintrich (1998) report that gender differences depend on the domain (e.g., boys are shown to be more intrinsically motivated in math and girls in languages).

[37]While this argument applies to extrinsic rewards in any form, monetary incentives are considered particularly insidious to intrinsic motivation.

[38]For further discussion see reviews by e.g., Eisenberger and Cameron 1996, Camerer and Hogarth 1999, Deci et al. 1999, Kohn 1999, Cameron and Pierce 2002. Frey and Oberholzer-Gee (1997) present a formal model and evidence from a field study of motivation crowding-out in an economic context.

several studies have tracked student performance after incentives are removed and generally find that students who received incentives continue to outperform the control group (see, e.g., Bettinger and Slonim 2007, Barrera-Osorio et al. 2008, Kremer et al. 2009, Levitt et al. 2010).[39]

We similalry explore whether the incentives have a detrimental impact on subsequent test performance. The richness of our design also permits us to learn whether spillovers differ between financial and non-financial incentives. Table 8 estimates the effects of incentives on performance in the same subject in the next testing period (Bloom and CPS) and in the subsequent subject taken in the same testing period (CPS).[40] Overall, we do not find strong evidence for crowding out – the point estimates are small in magnitude and generally not significant at conventional levels. We also do not find differences between financial and non-financial incentives.

Yet, there is some evidence for crowding out within the low financial incentives treatment. The point estimates are generally negative, with large and significant effects in Bloom. These results are consistent with the estimated effects of the $10 treatment on the incentivized test itself where the point estimates are also negative in Bloom and CPS (Table 3) as well as among several subgroups – older students (Table 4), girls (Table 7) and reading tests (Table 6) with the latter effects large and significant.

Hence, we do not find evidence for the wide-spread concern that paying kids once to perform well on a test has negative spillovers on future test performance. But in line with previous work, we do find some evidence that low level rewards can crowd out intrinsic motivation (Gneezy and Rustichini 2000a, 2000b).

[39]Additionally, Bettinger and Slonim (2007) find no evidence that a test performance incentive program erodes elementary school students' intrinsic motivation measured using student and teacher surveys.

[40]In Bloom (columns 1-2), we regress a student's winter 2009 treatment on her spring 2009 improvement (controlling for spring 2009 treatment). In CPS (columns 3-4), we regress a student's fall 2010 treatment on her winter 2011 improvement (controlling for winter 2011 treatment) and a student's winter 2011 treatment on her spring 2011 improvement (no treatments occurred in spring 2011). In CPS (columns 5-6), we regress math (reading) treatment on reading (math) improvement in the same period (fall 2010 or winter 2011) if a student received treatment on her first subject test. The covariates include additional controls for score on the treated test and (in column 6) baseline score in the subsequent subject.

5 Conclusion

This study uses a series of field experiments to explore various incentive schemes inspired by recent findings within behavioral economics. We explore the short-term effects of incentives on student effort and performance, varying the size and type of the rewards as well as their framing. We also offer rewards both immediately and with a delay. We find considerable promise for adding behavioral insights to the educational policymakers' toolkit. For example, we find evidence that incentives matter, and that their impact can be significantly enhanced if framed as a loss. Likewise, there is much to gain by broadening the scope of incentives to include both financial and non-financial variants.

Finally, the effect of timing of payoffs provides insights into the crux of the education problem that we face with our urban youth: effort is far removed from payout of rewards, making it difficult for students to connect them in a useful way. The failure to recognize this connection potentially leads to dramatic under-investment.

Continuing to apply important elements of behavioral economics to issues within education can directly aid practitioners in need of fresh solutions to the urban school problem. Such behavioral insights can also be used as a stepping stone for empiricists and experimentalists alike, who with the rich array of naturally-occurring data and experimental opportunities have a unique opportunity to examine theories heretofore untestable. Clearly, however, theory and empirical work must work symbiotically – there have been fewer theoretical advances that combine the best aspects of behavioral insights with issues germane to education. In this spirit, we hope that our study stimulates new work combining psychology and economics that not only deepens our understanding of empirical issues related to education, but also deepens our understanding of the important theoretical questions facing the field.

References

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

Angrist, Joshua D., Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics*, 1(1): 136-163.

Angrist, Joshua D., and Victor Lavy. 2009. "The Effect of High-Stakes High School Achievement Awards: Evidence from a Randomized Trial." *American Economic Review*, 99(4): 1384-1414.

Ariely, Dan, Anat Bracha, and Stephan Meier. 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review*. 99(1): 544–555.

Ashraf, Nava, Oriana Bandiera, and Kelsey Jack. 2011. "No margin, no mission? A field experiment on the role of incentives in the distribution of public goods." Working Paper.

Ball, Sheryl, Cathrine C. Eckel, Philip J. Grossman, and William Zame. 2001. "Status in Markets." *The Quarterly Journal of Economics*, 116(1): 161-188.

Barankay, Iwan. 2011. "Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment." Working Paper.

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 a Randomized Experiment in Colombia." NBER Working Paper No. 13890.

Baumert, Jürgen and Anke Demmrich. 2001. "Test motivation in the assessment of student skills: The effects of incentives on motivation and performance." *European Journal of Psychology of Education*, 16(3): 441-462.

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(1): 237-275.

Bénabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Eonomic Review*. 96(5): 1652-1678.

Berry, James. 2011. "Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India." Working Paper.

Bettinger, Eric. 2010. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." NBER Working Paper 16333.

Bettinger, Eric and Robert Slonim. 2007. "Patience among children." *Journal of Public Economics*, 91(1-2): 343-363.

Blanes i Vidal, Jordi and Mareike Nossol. 2011. "Tournaments Without Prizes: Evidence from Personnel Records." *Management Science*, 57(10): 1721-1736.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non. 2012. "Employee Recognition and Performance A Field Experiment." mimeo.

Braun, Henry, Irwin Kirsch, and Kentaro Yamamoto. 2011. "An Experimental Study of the Effects of Monetary Incentives on Performance on the 12th-Grade NAEP Reading Assessment." *Teachers College Record*.

Camerer, Colin F. and Robin M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty*, 19(1-3): 7-42.

Cameron, Judy and W. David Pierce. 2002. *Rewards and Intrinsic Motivation: Resolving the Controversy.* Bergin and Garvey: London.

Castillo, Marco, Paul Ferraro, Jeffrey L. Jordan, and Ragan Petrie. 2011. "The Today and Tomorrow of Kids: Time Preferences and Educational Outcomes of Children." *Journal of Public Economics*, 95(11-12): 1377-85.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, 125(6): 627-668.

Decker, Paul T., Daniel P. Mayer and Steven Galzerman. 2004. "The Effects of Teach for America on Students: Findings from a National Evaluation." Mathematica Policy Research Report No. 8792-750.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving," *The Quarterly Journal of Economics*, 127(1): 1-56.

Dobbie, Will and Roland G. Fryer. 2011. "Are High-Quality Schools Enough to Increase Achievment among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics*, 3(3): 158-87.

Eisenberger, Robert and Judy Cameron. 1996. "Detrimental effects of reward: Reality or myth?" *American Psychologist*, 51(11): 1153-1166.

Frey, Bruno S. 2007. "Awards as Compensation." *European Management Review*, 4: 6-14.

Frey, Bruno S. and Felix Oberholzer-Gee. 1997. "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out." *American Economic Review*, 87(4): 746-755.

Fryer, Roland G. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *The Quarterly Journal of Eonomics*, 126(4): 1755-1798.

Fryer, Roland G., Steven D. Levitt, John A. List, and Sally Sadoff. 2012. "Teacher Performance Pay." Working Paper.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics*, 118(3): 1049-1074.

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

Gneezy, Uri and Aldo Rustichini. 2000b. "A Fine is a Price." *Journal of Legal Studies*, 29(1): 1-17.

Gneezy, Uri and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics*, 118(3): 1049-1074.

Gneezy, Uri and Aldo Rustichini. 2004. "Gender and Competition at a Young Age." *American Economic Review*, 94(2): 377-381.
Gruber, Jonathan. 2001. Risky Behavior Among Youth: An Economic Analysis. Chicago: University of Chicago Press.

Hanemann, W. Michael. 1991. "Willigness To Pay and Willingness To Accept: How Much Can They Differ?" *American Economic Review*, 81(3): 635-647.

Hossain, Tanjim and John A. List. 2009. "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." NBER Working Paper 15623.

Huberman, Bernardo A., Christoph H. Loch and Ayse Onculer. 2004. "Status As a Valued Resource." *Social Psychology Quarterly*, 67(1): 103-114.

Jackson, C. Kirabo. 2010. "A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program." *Journal of Human Resources*, 45(3): 591-639.

Jacob, Brian. 2005. "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics*, 89(5-6): 761-796.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263-292.

Kirby, Kris N., Ricardo Godoy, Victoria Reyes-Garcia, Elizabeth Byron, Lilian Apaza, William Leonard, Eddy Perez, Vincent Vadez, and David Wilkie. 2002. "Correlates of delay-discount rates: Evidence from Tsimane' Amerindians of the Bolivian rain forest." *Journal of Economic Psychology*, 23(3): 291-316.

Kirby, Kris N., Gordon C. Winston and Mariana Santiesteban. 2005. "Impatience and Grades: Delay-Discount Rates Correlate Negatively with College GPA." *Learning and Individual Differences*, 15(3): 213-222.

Kohn, Alfie. 1999. *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. Boston: Houghton Mifflin.

Kosfeld, Michael and Susanne Neckermann. 2011. "Getting More Work for Nothing? Symbolic Awards and Worker Performance." *American Economic Journal: Microeconomics*, 3(3): 86-99.

Krawczyk, Michal. 2011. "To Answer or not to Answer? A Field Test of Loss Aversion." University of Warsaw Working Papers No. 13/2011 (53).

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

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

Laibson, David. 1997. "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics*, 112(2): 443-478.

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

Levitt, Steven D., John A. List and Sally Sadoff. 2010. "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment." Working Paper.

List, John A. and Anya C. Savikhin. 2012. "The Behavioralist as Dietician: Leveraging Behavioral Economics to Improve Child Food Choice and Consumption." Working Paper.

Loewenstein, George and Samuel Issacharoff. 1994. "Source Dependence in the Valuation of Objects." *Journal of Behavioral Decision Making*, 7(3): 157-168.

Mischel, Walter, Yuichi Shoda, and Monica L. Rodriguez. 1989. "Delay of Gratification in Children." *Science*, 244(4907): 933-938.

O'Donoghue, Ted and Matthew Rabin. 1999. "Doing it Now or Later." *American Economic Review*, 89(1): 103-124.

O'Neil, Harold F., Jr., Jamal Abedi, Charlotte Lee, Judy Miyoshi, and Ann Mastergeorge. 2004. "Monetary Incentives for Low-Stakes Tests." (CSE Rep. No. 625). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

O'Neil, Harold F., Jr., Brenda Sugrue, Eva L. Baker, and S. Golan. 1997. "Final report of experimental studies on motivation and NAEP test performance." (CSE Tech. Rep. No. 427). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Rivkin, Steven G., Eric A. Hanushe, and John F. Kain. 2005. "Teachers, Schools and Academic Achievement." *Econometrica*, 73(2): 417-458.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2): 247-252.

Samuelson, William and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty*, 1(1): 7-59.

Schultz, T. Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." *Journal of Development Economics*, 74(1): 199-250.

Skaalvik, Sidsel and Einar M. Skaalvik. 2005. "Self-Concept, Motivational Orientation, and Help-Seeking Behavior in Mathematics: A Study of Adults Returning to High School." *Social Psychology of Education*, 8(3): 285-302.

Steinberg, Laurence,Sandra Graham, Lia O'Brien, Jennifer Woolard, Elizabeth Cauffman, and Marie Banich. 2009. "Age Differences in Future Orientation and Delay Discounting." *Child Development*, 80(1): 28-44.

Strotz, Robert H. 1955. "Myopia and Inconsistency in Dynamic Utility Maximization" *The Review of Economic Studies*, 23(3): 165-180.

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice" *Journal of Economic Behavior & Organization*, 1(1): 39-60.

Tran, Anh and Richard Zeckhauser. 2009. "Rank as an Incentive." Harvard Working Paper RWP09-019.

Tversky, Amos and Daniel Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *The Quarterly Journal of Economics*, 106(4): 1039-1061.

Vallerand, Robert J., Luc G. Pelletier, Marc R. Blais, Nathalie M. Brière, Caroline Senécal, and Evelyne F. Vallières. 1992. "The Academic Motivation Scale: A Measure of Intrinsic, Extrinsic, and Amotivation in Education." *Educational and Psychological Measurement*, 52(4): 1003-1017.

Vallerand, Robert J., Françoys Gagné, Caroline Senécal, and Luc G. Pelletier. 1994. "A Comparison of the School Intrinsic Motivation and Perceived Competence of Gifted and Regular Students." *Gifted Child Quarterly*, 38(4): 172-175.

Waldfogel, Joel. 1993. "The Deadweight Loss of Christmas." *American Economic Review*, 83(5): 1328-1336.

Wolters, Christopher A. and Paul R. Pintrich. 1998. "Contexual Differences in Student Motivation and Self-regulated Learning in Mathematics, English, and Social Studies Classrooms." *Instructional Science*, 26(1-2): 27-47.

Table 1: Overview of the Treatments

	Bloom High School 2009	Bloom High School 2010	CH Elementary 2010	CPS Elementary 2010	CPS Elementary 2011
Control - No statement	X				X
Control - Statement		X	X[a]	X[b]	X
Financial Low ($10)	X		X[a]	X	
Financial High ($20)	X	X	X	X	X
Non-Financial (Trophy)			X	X	X
Financial Loss		X		X	X
Non-Financial Loss				X	X
Financial Delayed				X	
Non-Financial Delayed				X	
Financial Loss Delayed				X	
Non-Financial Loss Delayed				X	
Test subject - Reading	X	X		X	X
Test subject - Math			X	X	X

Note: Financial Loss, Financial Delayed and Financial Loss Delayed all received Financial High ($20) incentives. Non-Financial Loss, Non-Financial Delayed and Non-Financial Loss Delayed all received Non-Financial (trophy) incentives.

[a] Control and Financial Low ($10) are each pooled with "Comparison" treatments that add a statement that a student's improvement will be compared to three other students with similar past scores (see Appendix A for scripts). The comparison statement did not significantly affect test performance at the 10% level.

[b] Control - Statement is pooled with Control - Statement - Delayed which states that students will learn their scores "one month after the test" instead of "immediately after the test" (see Appendix A for scripts). The delayed statement did not significantly affect test performance at the 10% level.

Table 2a: Baseline Characteristics by Treatment Group: Bloom High School

	Control	Financial Low	Financial High	Financial Loss
N	285	166	324	154
Baseline Score	0.112	0.086	-0.070	0.174
	(0.954)	(0.900)	(0.956)	(1.040)
Female	0.523	0.524	0.435	0.468
	(0.500)	(0.501)	(0.497)	(0.501)
Black	0.586	0.452**	0.556	0.468
	(0.493)	(0.499)	(0.498)	(0.501)
Hispanic	0.288	0.422**	0.306	0.318
	(0.453)	(0.495)	(0.461)	(0.467)
Free/Reduced Lunch	0.716	0.711	0.701	0.740
	(0.452)	(0.455)	(0.459)	(0.440)

Note: The table reports group means (baseline score) and proportions (female, black, Hispanic, free/reduced lunch) pooling the Bloom 2009 and Bloom 2010 waves. Standard deviations are reported in parentheses. Baseline score is standardized within wave using the full sample of Bloom students. Asterisks indicate a difference of means/proportions (compared to pooled control with standard errors clustered by class) significant at the 10/5/1 percent level.

Table 2b: Baseline Characteristics by Treatment Group: CH Elementary

	Control	Financial Low	Financial High	Non-Financial
N	179	165	30	69
Baseline Score	-0.511	-0.510	-0.399	-0.682
	(0.765)	(0.781)	(1.067)	(0.775)
Grade	6.179	5.133***	5.400**	5.072***
	(1.958)	(1.446)	(1.380)	(1.229)
Female	0.503	0.497	0.433	0.449
	(0.501)	(0.502)	(0.504)	(0.501)
Black	0.497	0.461	0.300**	0.290***
	(0.501)	(0.500)	(0.466)	(0.457)
Hispanic	0.391	0.461	0.633**	0.623***
	(0.489)	(0.500)	(0.490)	(0.488)
Free/Reduced Lunch	0.877	0.891	0.900	0.928
	(0.329)	(0.313)	(0.305)	(0.261)
IEP	0.109	0.070	0.034	0.101
	(0.313)	(0.257)	(0.186)	(0.304)

Note: The table reports group means (baseline score, grade) and proportions (female, black, Hispanic, free/reduced lunch, IEP). Standard deviations are reported in parentheses. Baseline score is standardized within grade using the full sample of Illinois students. Asterisks indicate a difference of means/proportions (compared to pooled control with standard errors clustered by school-grade) significant at the 10/5/1 percent level.

Table 2c: Baseline Characteristics by Treatment Group: CPS Elementary

	Control	Financial Low	Financial High	Non-Financial	Financial Loss	Non-Financial Loss	Financial Delayed	Non-Financial Delayed	Financial Loss Delayed	Non-Financial Loss Delayed
N	3123	108	666	556	757	730	53	77	30	82
Baseline Score	-0.417	-0.323	-0.302	-0.247	-0.361	-0.394	-0.263	-0.506	-0.590	-0.244
	(0.913)	(0.926)	(0.895)	(0.897)	(0.872)	(0.895)	(0.908)	(0.877)	(0.948)	(0.906)
Grade	5.090	5.648	5.335	5.207	4.786	5.027	5.547	4.584	4.833	4.756
	(1.867)	(1.362)	(1.727)	(1.968)	(1.922)	(1.883)	(0.503)	(1.641)	(1.599)	(2.016)
Test Subject - Math	0.306	0.231*	0.254***	0.185***	0.252***	0.351**	0.604***	0.208*	0.000***	0.000***
	(0.461)	(0.424)	(0.435)	(0.389)	(0.435)	(0.478)	(0.494)	(0.408)	(0.000)	(0.000)
Female	0.481	0.676***	0.535**	0.504	0.555***	0.470	0.472	0.416	0.367	0.488
	(0.500)	(0.470)	(0.499)	(0.500)	(0.497)	(0.499)	(0.504)	(0.496)	(0.490)	(0.503)
Black	0.932	0.926	0.953**	0.982***	0.951*	0.959***	0.943	0.909	0.933	0.915
	(0.252)	(0.263)	(0.211)	(0.133)	(0.216)	(0.199)	(0.233)	(0.289)	(0.254)	(0.281)
Free/Reduced Lunch	0.905	0.944	0.949***	0.908	0.863***	0.892	0.925	0.623***	0.833	0.500***
	(0.293)	(0.230)	(0.220)	(0.289)	(0.344)	(0.311)	(0.267)	(0.488)	(0.379)	(0.503)
IEP	0.091	0.083	0.099	0.090	0.102	0.096	0.038	0.104	0.067	0.037*
	(0.287)	(0.278)	(0.299)	(0.286)	(0.302)	(0.295)	(0.192)	(0.307)	(0.254)	(0.189)

Note: The table reports group means (baseline score, grade) and proportions (test subject - math, female, black, Hispanic, free/reduced lunch, IEP) pooling CPS 2010 and CPS 2011 waves. Standard deviations are reported in parentheses. Baseline score is standardized within wave, test subject and grade using the full sample of CPS students. Asterisks indicate a difference of means/proportions (compared to pooled control with standard errors clustered by school-grade) significant at the 10/5/1 percent level.

Table 3: Treatment Effects on Test Score Improvement

	Bloom High School		CH Elementary		CPS Elementary	
Financial Low	-0.028	-0.069	0.205**	0.237**	-0.065	-0.028
	(0.065)	(0.053)	(0.097)	(0.088)	(0.110)	(0.056)
Financial High	0.195***	0.120**	0.456***	0.373*	0.057	0.089**
	(0.051)	(0.052)	(0.122)	(0.190)	(0.047)	(0.040)
Non-Financial			0.317***	0.262**	0.010	0.063
			(0.095)	(0.102)	(0.045)	(0.039)
Financial Loss	0.172***	0.148**			0.083**	0.103***
	(0.051)	(0.071)			(0.039)	(0.036)
Non-Financial Loss					0.078*	0.119***
					(0.045)	(0.045)
Session	Yes	Yes			Yes	Yes
Other Covariates		Yes		Yes		Yes
Students	825	825	423	423	5577	5577
Classes/School-Grades	38	38	21	21	167	167

Note: The table reports OLS estimates for treatment effects on test score improvement in standard deviation units for pooled sessions in Bloom and CPS and a single session in CH. Robust standard errors clustered by class in Bloom and by school & grade in CH and CPS are reported in parentheses. The omitted category in each regression is the pooled control group. Columns (1)-(2) and (4)-(5) control for session. Columns (2), (4) and (6) add controls for baseline score on the tested subject (score, score squared and score cubed), past treatment (Bloom and CPS), school (CH and CPS), grade (CH and CPS), teacher (Bloom), test subject (CPS) and demographics (gender, race/ethnicity,language, free/reduced lunch status and (in CH and CPS) IEP status). Asterisks indicate significance at the 10/5/1 percent level.

Table 4: Treatment Effects by Grade

	CPS			
	2nd-5th Grade		6th-8th Grade	
Financial Low	0.153***	0.011	-0.110**	0.002
	(0.054)	(0.117)	(0.042)	(0.080)
Financial High	0.121*	0.140**	-0.018	-0.009
	(0.067)	(0.056)	(0.054)	(0.056)
Non-Financial	0.055	0.122**	-0.060	0.038
	(0.059)	(0.061)	(0.060)	(0.065)
Financial Loss	0.059	0.098*	0.130**	0.117**
	(0.054)	(0.051)	(0.051)	(0.046)
Non-Financial Loss	0.176***	0.249***	-0.101**	-0.012
	(0.048)	(0.057)	(0.048)	(0.052)
Session	Yes	Yes	Yes	Yes
Other Covariates		Yes		Yes
Students	3203	3203	2374	2374
School-Grades	112	112	73	73

Note: The table reports OLS estimates for treatment effects on test score improvement in standard deviation units for pooled sessions. Robust standard errors clustered by school & grade are reported in parentheses. The omitted category in each regression is the pooled control group. All regressions control for session. Columns (2) and (4) add controls for baseline score on the tested subject (score, score squared and score cubed), past treatment, school, grade, test subject and demographics (gender, race/ethnicity, language, free/reduced lunch status and IEP status). Asterisks indicate significance at the 10/5/1 percent level.

Table 5: Effect of Delayed Rewards

	CPS			
	Immediate		Delayed	
Financial High	0.057	0.089**	-0.189	-0.048
	(0.047)	(0.040)	(0.159)	(0.163)
Non-Financial	0.010	0.065*	-0.164	-0.085
	(0.045)	(0.039)	(0.118)	(0.075)
Financial Loss	0.083**	0.102***	-0.048	0.198
	(0.039)	(0.036)	(0.156)	(0.131)
Non-Financial Loss	0.078*	0.120***	-0.160	-0.105
	(0.045)	(0.045)	(0.121)	(0.126)
Session	Yes	Yes		
Other Covariates		Yes		Yes
Students	5471	5471	542	542
School-Grades	167	167	38	38

Note: The table reports OLS estimates for treatment effects on test score improvement in standard deviation units for pooled sessions in columns (1)-(2) and a single session in columns (3)-(4) (delayed treatments occurred in only one session). Robust standard errors clustered by school & grade are reported in parentheses. The omitted category in each regression is the pooled control group for the relevant session(s). Columns (1)-(2) control for session. Column (3) contains no controls. Columns (2) and (4) add controls for baseline score on the tested subject (score, score squared and score cubed), past treatment, school, grade, test subject and demographics (gender, race/ethnicity, language, free/reduced lunch status and IEP status). Asterisks indicate significance at the 10/5/1 percent level.

Table 6: Treatment Effects by Test Subject

	CPS			
	Reading		Math	
Financial Low	-0.120**	-0.142***	0.100	0.186
	(0.058)	(0.051)	(0.063)	(0.245)
Financial High	0.024	0.072*	0.198**	0.307***
	(0.054)	(0.041)	(0.093)	(0.077)
Non-Financial	0.001	0.053	0.081	0.169
	(0.050)	(0.048)	(0.081)	(0.112)
Financial Loss	0.038	0.036	0.230***	0.353***
	(0.038)	(0.035)	(0.069)	(0.101)
Non-Financial Loss	0.089	0.122**	0.078	0.171**
	(0.057)	(0.057)	(0.078)	(0.079)
Session	Yes	Yes	Yes	Yes
Other covariates		Yes		Yes
Students	3953	3953	1605	1605
School-Grades	131	131	74	74

Note: The table reports OLS estimates for treatment effects on test score improvement in standard deviation units for pooled sessions. Robust standard errors clustered by school & grade are reported in parentheses. The omitted category in each regression is the pooled control group. All regressions control for session. Columns (2) and (4) add controls for baseline score on the tested subject (score, score squared and score cubed), past treatment, school, grade and demographics (gender, race/ethnicity, language, free/reduced lunch status and IEP status). Asterisks indicate significance at the 10/5/1 percent level.

Table 7: Treatment Effects by Gender

	Bloom				CPS			
	Female		Male		Female		Male	
Financial Low	-0.191*	-0.161**	0.160*	0.065	-0.085	0.000	-0.052	-0.077
	(0.105)	(0.077)	(0.092)	(0.099)	(0.097)	(0.072)	(0.123)	(0.081)
Financial High	0.164**	0.110	0.226***	0.143*	0.023	0.070	0.097*	0.113**
	(0.066)	(0.072)	(0.065)	(0.076)	(0.052)	(0.044)	(0.057)	(0.055)
Non-Financial					-0.030	0.019	0.048	0.117**
					(0.054)	(0.050)	(0.049)	(0.047)
Financial Loss	0.080	0.031	0.225***	0.257**	0.051	0.061	0.116**	0.159***
	(0.080)	(0.096)	(0.081)	(0.101)	(0.043)	(0.043)	(0.051)	(0.052)
Non-Financial Loss					0.052	0.086*	0.106*	0.152***
					(0.050)	(0.049)	(0.054)	(0.055)
Session	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Covariates		Yes		Yes		Yes		Yes
Students	412	412	413	413	2829	2829	2748	2748
Classes/School-Grades	38	38	38	38	167	167	166	166

Note: The table reports OLS estimates for treatment effects on test score improvement in standard deviation units for pooled sessions. Robust standard errors clustered by class in Bloom and by school & grade in CPS are reported in parentheses. The omitted category in each regression is the pooled control group. All regressions control for session. Columns (2), (4), (6) and (8) add controls for baseline score on the tested subject (score, score squared and score cubed), past treatment (CPS), school (CPS), grade (CPS), teacher (Bloom), test subject (CPS) and demographics (gender, race/ethnicity, language, free/reduced lunch status and (in CPS) IEP status). Asterisks indicate significance at the 10/5/1 percent level.

Table 8: Treatment Effects on Future Tests

	Bloom		CPS			
	Same Subject Subsequent Test Session		Same Subject Subsequent Test Session		Subsequent Subject Same Test Session	
Financial Low	-0.324***	-0.295*	0.014	-0.023	-0.115*	0.046
	(0.098)	(0.137)	(0.090)	(0.082)	(0.064)	(0.077)
Financial High	0.012	-0.045	-0.065*	-0.021	0.027	-0.065
	(0.130)	(0.157)	(0.034)	(0.037)	(0.063)	(0.043)
Non-Financial			-0.035	-0.040	0.042	-0.022
			(0.035)	(0.032)	(0.086)	(0.052)
Financial Loss			-0.035	-0.040	0.026	-0.011
			(0.033)	(0.027)	(0.074)	(0.042)
Non-Financial Loss			-0.038	-0.005	0.090	0.054
			(0.040)	(0.037)	(0.082)	(0.043)
Subsequent Test Treatment	Yes	Yes	Yes	Yes		
Session			Yes	Yes	Yes	Yes
Other Covariates		Yes		Yes		Yes
Students	268	268	5298	5298	4574	4574
Classes/School-Grades	13	13	166	166	160	160

Note: The table reports OLS estimates for treatment effects on test score improvement in standard deviation units for pooled sessions in CPS and a single session in Bloom (a Subsequent Test Session occurred in only one session in Bloom). Robust standard errors clustered by class in Bloom and by school & grade in CPS are reported in parentheses. Columns (1) - (4) control for treatment on the subsequent test. Columns (3) - (6) include controls for session. Columns (2), (4) and (6) add controls for baseline score on the tested subject (score, score squared and score cubed), past treatment (CPS), school (CPS), grade (CPS), test subject (CPS) and demographics (gender, race/ethnicity, language, free/reduced lunch status and (in CPS) IEP status). Asterisks indicate significance at the 10/5/1 percent level.

A Appendix: Administrator Scripts

A.1 Bloom

Common to all treatments

To the teacher:

Please read the following statement to your students immediately before they begin the STAR test (after you have given them your regular instructions for testing):

Bloom 2009

Financial Low ($10) You are about to take the STAR Reading Assessment. You also took the STAR Reading Assessment in the fall. If your score on the STAR today is higher than your score in the fall, you will receive $10. You will be paid at the end of the test. Please fill out your name, signature and date on the assent form. You will turn this in at the end of the test.

Financial High ($20) You are about to take the STAR Reading Assessment. You also took the STAR Reading Assessment in the fall. If your score on the STAR today is higher than your score in the fall, you will receive $20. You will be paid at the end of the test. Please fill out your name, signature and date on the assent form. You will turn this in at the end of the test.

Bloom 2010

Control - Statement

You are about to take the STAR Reading Assessment. You also took the STAR Reading Assessment in the fall. Please try to improve your score from the fall.

Financial High ($20) You are about to take the STAR Reading Assessment. You also

took the STAR Reading Assessment in the fall. If your score on the STAR today is higher than your score in the fall, you will receive $20. You will be paid at the end of the test. Please fill out your name, signature and date on the assent form. You will turn this in at the end of the test.

Financial Loss ($20) You are about to take the STAR Reading Assessment. You also took the STAR Reading Assessment in the fall. Please try to improve your score from the fall.
In front of you is an envelope that contains $20. Please open the envelope to confirm that there is $20 inside. [*Wait for students to open envelope and sign confirmation form.*]
If you improve your score from the fall, you will get to keep the $20. If you do not improve your score from the fall, you will not get to keep the $20. You will have to return the $20 immediately after the test.

A.2 Chicago Heights

Common to all treatments

To the teacher:
Please read the following statement to your students immediately before they begin the STAR test (after you have given them your regular instructions for testing):

Control - Statement

You are about to take the ThinkLink Learning test. You also took ThinkLink in the winter. Please try to improve your score from the winter.

Control - Statement - Comparison

You are about to take the ThinkLink Learning test. You also took ThinkLink in the winter. Please try to improve your score from the winter. We will compare your improvement to 3

other students who had the same score as you in the winter.

Financial Low ($10)

You are about to take the ThinkLink Learning test. You also took ThinkLink in the winter. Please try to improve your score from the winter. If you improve your score from the winter, you will receive $10. You will be paid in cash immediately after the test.

Financial Low ($10) - Comparison

You are about to take the ThinkLink Learning test. You also took ThinkLink in the winter. Please try to improve your score from the winter. We will compare your improvement to 3 other students who had the same score as you in the winter. If you improve your score from the winter, you will receive $10. You will be paid in cash immediately after the test.

Financial High ($20)

You are about to take the ThinkLink Learning test. You also took ThinkLink in the winter. Please try to improve your score from the winter. If you improve your score from the winter, you will receive $20. You will be paid in cash immediately after the test.

Non-Financial (Trophy)

You are about to take the ThinkLink Learning test. You also took ThinkLink in the winter. Please try to improve your score from the winter. If you improve your score from the winter, you will receive this trophy and we will post a photo like this of you in the class [show sample photo]. You will receive the trophy and be photographed immediately after the test.

A.3 Chicago Public School

Common to all treatments

To the teacher:

Please read the following statement to your students immediately before they begin the Scantron test (after you have given them your regular instructions for testing):

Control - Statement

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring. You will learn your score immediately after the test.

Control - Statement - Delayed

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring. You will learn your score one month after the test.

Financial Low ($10)

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring. If you improve your score from the spring, you will receive $10. You will learn your score and be paid in cash immediately after the test.

Financial High ($20)

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring. If you improve your score from the spring, you will receive $20. You will learn your score and be paid in cash immediately after the test.

Financial High ($20) - Delayed

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring. If you improve your score from the spring, you will

receive $20. You will learn your score and be paid in cash one month after the test.

Financial Loss ($20)

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring.

You are being given an envelope that contains $20. Please open the envelope to make sure that there is $20 inside. Please sign the form that says that this is your $20. And write down what you will do with your $20. [*Wait for students to open envelope and complete the confirmation form.*]

If you improve your score from the spring, you will get to keep your $20. If you do not improve your score from the spring, you will have to return your $20. You will learn your score and whether you get to keep your $20 immediately after the test

Financial Loss ($20) - Delayed

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring.

You are being given an envelope that contains $20. Please open the envelope to make sure that there is $20 inside. Please sign the form that says that this is your $20. And write down what you will do with your $20.[*Wait for students to open envelope and complete the confirmation form.*]

If you improve your score from the spring, you will get to keep your $20. If you do not improve your score from the spring, you will have to return your $20. You will learn your score and whether you get to keep your $20 one month after the test.

Non-Financial (Trophy)

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring. If you improve your score from the spring, you will

receive this trophy [*SHOW SAMPLE TROPHY*]. You will learn your score and receive the trophy immediately after the test.

Non-Financial (Trophy) - Delayed

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring. If you improve your score from the spring, you will receive this trophy [*SHOW SAMPLE TROPHY*]. You will learn your score and receive the trophy one month after the test.

Non-Financial Loss (Trophy)

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring.

You are being given a trophy. Please sign the form that says that this is your trophy. And write down what you will do with your trophy. [*Wait for students to complete the confirmation form.*]

If you improve your score from the spring, you will get to keep the trophy [*SHOW SAMPLE TROPHY*]. If you do not improve your score from the spring, you will have to return your trophy. You will learn your score and whether you get to keep your trophy immediately after the test.

Non-Financial Loss (Trophy) - Delayed

You are about to take the Scantron test. You also took Scantron in the spring. Please try to improve your score from the spring.

You are being given a trophy. Please sign the form that says that this is your trophy. And write down what you will do with your trophy. [*Wait for students to complete the confirmation form.*]

If you improve your score from the spring, you will get to keep the trophy [*SHOW SAMPLE*

TROPHY]. If you do not improve your score from the spring, you will have to return your trophy. You will learn your score and whether you get to keep your trophy one month after the test.