The Impact of Short-term Incentives on Student Performance*

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

September, 2011

Abstract

Although direct financial incentives are prevalent in nearly all walks of life, they are rarely used on students. Most prior studies of incentives have focused on interventions that operate over a time horizon of months or years, typically yielding small positive impacts. In this paper, we report on the results of immediate and transitory incentives implemented among elementary and high school students in three school districts. We announce the incentives immediately before a standardized test, isolating the role of short-term effort in achievement. We also vary both the type of incentives (financial and non-financial) and their framing (as gains and as losses). We find that incentives affect student performance, although there is substantial variation across settings. Incentives framed as losses have consistently large effects relative to other educational interventions ($0.12 - 0.22$ standard deviations). We find mixed evidence on the impact of incentives framed as gains with large effects in two school districts ($0.2 - 0.4$ 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 uncover a variety of determinants such as age, gender, and test subject that influence reward effectiveness. All motivating power 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. Our findings also imply that in the absence of immediate incentives,

*Affiliations: Levitt and List: University of Chicago and NBER; Neckermann: ZEW & University of Mannheim; Sadoff: University of California San Diego. 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, Ian Muir, Joseph Seidel and Phuong Ta 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.

> 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.

Financial incentives are ubiquitous in modern society. Traditionally, however, direct financial incentives have not 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.[1] 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 (see e.g., Bettinger and Slonim 2007, Steinberg et al 2009 for further discussion). They may therefore respond more strongly to rewards with very short time horizons compared to incentives extending over several months or years.

In light of the modest effectiveness of financial rewards and a general uneasiness among educators with using such incentives, educators and policy makers are looking for alternative and more cost effective means to increase student effort. Insights from behavioral economics can be useful for the educational context in this regard. One potentially fruitful avenue is non-financial incentives. Recent advances in behavioral economics have shown that 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 motivating power from a variety of mechanisms such as status, self-image concerns, and relative perfor-

[1]Examples include Progresa in Mexico which offered offered incentives for school enrollment and attendance (Schultz 2004, Behrman, Sengupta, and Todd 2005). A similar conditional cash transfer program was instituted in Colombia (Barrera-Osorio, Bertrand, Linden and Perez-Calle 2008). Other programs have based rewards on overall school performance (see Angrist, Bettinger and Kremer 2006, Leuven, Oosterbeek and vander Klaauw forthcoming, Levitt, List and Sadoff 2010). Fryer (2010) reports on a series of incentive programs carried out in a number of large American school districts.

mance feedback that have been shown to affect behavior.[2] Another avenue is framing, i.e. the manner in which a decision is presented. In a labor context, Hossain and List (2009) framed conditional incentives as "losses" or "gains" and show that the former elicit significantly higher effort. Such effects are closely related to other 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 (Hanneman, 1991). They are broadly consistent with a notion of loss aversion, an insight gained from Kahneman and Tversky's (1979) prospect theory, which surmises that changes relative to a neutral reference point affect utility more than absolute levels and that losses loom larger than gains.

In this work, we investigate the immediate impact of incentives on student performance on otherwise low-stakes standardized tests. Between January 2009 and January 2011, we conducted randomized field experiments among elementary and high school students in three school districts in and around Chicago. Students were offered cash or non-pecuniary rewards for an improvement in test scores. We investigate the effectiveness of low and high financial incentives ($10, $20) and compare this to the impact of non-financial incentives in the form of a trophy for achievement. These incentives were presented in either the gain or in the loss domain. The tests lasted between 15 - 60 minutes, yielding a high hourly wage (particularly in the highest financial incentive group) that we expect is salient among our subject pool of low-income children and adolescents.

The incentives were announced immediately before the test (with no advance notice). This design allows us to estimate the effect of incentives on performance solely through inducement of greater short-run effort – it avoids confounding due to discount rates or human capital accumulation (i.e., studying for the test). In addition, we directly test the importance of discount rates by offering incentives either immediately after the test ended

[2]See Huberman et al., 2004 and Ball et al., 2001 on status; Barankay, 2011, Tran and Zeckhauser, 2010, Blanes i Vidal and Nossol, 2009 on relative performance feedback; and Ariely et al., 2009 and Della Vigna et al., 2010 on image motivation and social pressure.

or with a delay (a month after the test).

Most previous programs that 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 2007); and standardized tests for elementary/middle school students in Kenya (Kremer, Miguel and Thornton, forthcoming), India (Berry 2008), Ohio (Bettinger 2008) 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. ONeil et al (1992, 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).[3] In a similar design, Baumer 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.

As far as we know, Braun, Kirsch and Yamamoto (2011), which we became aware of only after our field work was completed, 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 $35 to eighth and twelfth graders on a low stakes 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. As far as we know, no previous studies have tested immediate rewards that vary the level of the performance-based incentive, and compare financial to non-financial incentives in a single setting. Moreover, we are not aware of studies that have explicitly studied the relative effectiveness of rewards framed as either gains or losses in the educational context.[4]

[3]ONeil (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.

[4]In the context of education, Krawacyzyk (2011) tests the effect of framing on risk taking on a final exam

This research was undertaken with a total of approximately 6,500 students in a variety of different settings in order to to explicitly investigate the robustness of results in terms of their replicability (different elementary schools) as well as their transferability to other settings (elementary school, high school) as well as the possibility to scale up incentives schemes from smaller to larger districts. Most previous studies report findings from a single experiment without any attempt to replicate their results in a similar set-up to address the transferability to different settings and different scales. This paper addresses both questions by studying the impact of incentives in different settings, with students of different ages and in school districts of different sizes.[5]

We find that incentives affect student performance, although there is substantial variation in the effectiveness of rewards across settings. We find that incentives framed as losses have consistently large effects relative to other educational interventions ($0.12 - 0.22$ standard deviations). We find mixed evidence on the impact of incentives framed as gains with large effects in two school districts ($0.2 - 0.4$ 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. Effects are more pronounced for Math tests than for Reading.

Our results suggest that in the absence of immediate incentives, many students put forth low effort on standardized tests. The findings have important implications for schools

and finds no effect. The study does not examine the effect of framing on effort or overall performance.

[5]In a similar vein, Braun et al (2011) include 2,600 students in 59 schools in seven states. Fryer (2011) tests various incentive designs in multiple urban school districts.

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 also create biases in measures of student ability, teacher value added, school quality and achievement gaps.[6] 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 could induce sub-optimal effort.

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

2 Experimental Design & 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 7 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.

The 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

[6]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.

external reason to do well on. The scores are not used as a measure of student or teacher performance – in fact, many students never learn their score, as they are used primarily as a diagnostic tool for teachers to help prepare students for the high-stakes standardized state test at the end of the year. Students generally take the tests three times a year in the fall, winter and spring.[7] They are computer-based and last between 15-60 minutes with students' results available immediately after the test ends.[8]

In each session, immediately before testing began, the test administrator announced the incentive and told students they would receive the reward immediately after the test ended if they improved upon their baseline score from a prior testing session.[9] Immediately after the test ended, we handed out rewards to qualifiying students, except in the case of delayed rewards which were distributed a month after testing.[10]. 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 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.[11] The test administrator did not make any announcement in control (no statement) groups. In control (statement) groups, the administrator encouraged students to improve on the test but did not offer any

[7]In Chicago Heights, students also take a baseline test at the beginning of the year.

[8]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 last about 30 minutes. In CPS, the experiment took place during either the math or reading portion of the Scantron Performance Series, which each lasts about 60 minutes.

[9]The test administrator was either the students' teacher or a literacy coach at the school. He or she announced the incentive immediately after providing students with instructions on how to take the test. Our Research Assistants were present in the room while the script was read to the children and handed out the rewards in the loss treatments. Then they waited in a room next door to the testing room while testing took place and returned to distribute the rewards after all testing had ended.

[10]In CPS, some 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

[11]In addition to framing and loss aversion, this condition might affect behavior as the reward is more salient - sitting right in front of the student - and increases trust and subjective beliefs with respect to actual payout of these unusual incentives.

incentive to do so (this allows us to isolate the incentive effect from any effect caused by the mere request to improve). The scripts for the different treatments can be found in Appendix A. An overview of the treatments conducted are presented in Table 1.[12]

We randomized at the level of English class (Bloom) or school-grade (CH and CPS) and blocked the randomization on average baseline score, grade and race/ethnicity, when available.[13] In cases where students participated in two waves (Bloom 2009 and CPS 2010/2011), we re-randomized for the second wave. Thus, some students received the same treatment in both sessions, while others received a different incentive in the two sessions. Students received no advance notice of the incentives prior to the testing sessions.[14] In the two cases, where students had 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.[15]

[12]The various waves included additional incentives not reported here (this paper only includes the incentives common across the three settings). Results on the other treatments, as well as on the longer term effects of the rewards, will be reported in follow-up papers.

[13]In Bloom, we blocked on baseline test score. If the baseline score was not available, we blocked classes by their track: regular, remedial or honors. In Chicago Heights, we blocked on baseline test score and race/ethnicity. In CPS, we blocked on school, baseline test score, grade, and past treatment (in the second wave).

[14]One week before testing, we sent home a consent form to parents stating that, we would like their child 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 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. No 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 treatment and 71% of forms were returned prior to the winter treatment. In order to participate, students 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.

[15]An analysis of potential spillover effects will be addressed in a follow-up paper. 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.

3 Results

Tables 2-4 reports the sample means by treatment group for pre-treatment characteristics in Bloom (2009 and 2010), Chicago Heights (2010) and CPS (2010 and 2011), with standard errors clustered by class (Bloom) or school-grade (CH and CPS).[16] The tables indicate the presence of some significant differences between treatment and control group means. In Bloom (Table 2), the only significant differences are the proportion of African-American and Hispanic students in the financial low ($10) treatment. In Chicago Heights (Table 3), all three treatment groups have a significantly lower average grade than control. Financial high and non-financial have significantly lower percentages of African-American students and significantly higher proportions of Hispanic students than control. In CPS, (Table 4), the various treatment groups are balanced on baseline score (the non-financial incentive group has higher baseline scores than control significant at the $p < 0.1$ level). The proportion of females, African-Americans and students eligible for free/reduced lunch are statistically significantly different in some instances. As shown below, including controls for pre-treatment characteristics as well as baseline performance does not alter the results.

Tables 5-7 present estimated treatment effects on test score improvement in Bloom, Chicago Heights and CPS. The dependent variable is test score improvement (in standard deviation units).[17] The first column reports treatment effect estimates absent any controls except for the wave (session) of the experiment (Bloom and CPS). The second column adds

[16]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 CH 2010, winter 2010 serves as the baseline. In CPS 2010, spring 2010 serves as the baseline. In CPS 2011, fall 2010 serves as the baseline. Note that for some students, their fall 2010 score was incentivized in the previous wave of the experiment. Excluding students with previously incentivized scores does not affect the results.

[17]Improvement is measured as the difference between the standardized outcome score and the standardized baseline score. Scores are standardized to have mean zero and standard deviation equal to 1. In Bloom, we standardize scores within each test session (fall 2008, winter 2009, spring 2009, fall 2009 and spring 2010) using the full sample of Bloom students in baseline sessions (fall 2008 and 2009) and the control group in experimental sessions (winter 2009, spring 2009 and 2010). In Chicago Heights, we standardize scores within each test session (winter and spring 2010) and grade using the full sample of Illinois students. In CPS, we standardize scores within test session (spring 2010, fall 2010 and winter 2011), test subject and grade using the full population of CPS students.

controls for baseline fall score (score, score squared and score cubed) and past treatment.[18] Column three adds controls for test subject (CPS), school and grade (CH and CPS), and teacher fixed effects (Bloom). The final column adds demographics controls for gender, race/ethnicity and free/reduced lunch status.

Table 5 reports the effects of financial incentives on test score improvement among Bloom sophomores. We pool the three waves of the study (winter 2009, spring 2009 and spring 2010).[19] The estimates for the effect of the $20 incentives (financial high and financial loss) are all positive and statistically significant. The magnitude of the impact of the $20 incentives is substantively large: an increase in test scores of roughly 0.2 standard deviations. This increase is equivalent to about 5-6 months worth of learning.[20] The estimated coefficients for the low financial incentive ($10) suggests that is has little or no effect. The $20 incentive framed as a loss has the same impact on performance as $20 framed as a gain (they are both significantly different from the $10 incentive at the $p < .05$ level).

The main results from Chicago Heights are presented in Table 6.[21] All incentives have a positive and statistically significant impact on improvement and the coefficients are robust over the different specifications of the model. Low financial incentives ($10) increase improvement by $0.18 - 0.24$ standard deviations. High financial incentives ($20) increase improvement by $0.38 - 0.46$ standard deviations. The point estimates for the non-financial incentives ($0.24 - 0.32$ standard deviations) fall between those for the $10 and $20 treat-

[18]Past treatment controls for whether the person was incentivized in the previous wave. In a follow-up paper, we will look at such spill-over effects in more detail.

[19]The pooled control group consists of the 2009 control group (Control - No statement) and the 2010 control group (Control - Statement). There are no statistically significant differences between the two control groups and pooling does not affect the results. An explicit comparison of the two different control groups is incorporated into the design of the CPS experiment and is reported below. An analysis of the individual waves separately yields similar results.

[20]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).

[21]The Control and $10 groups are each pooled with different, but very similar treatments that add the statement that a student's improvement will be compared to 3 other students with similar past scores (scripts denote with "comparison" in section A). This comparison statement did not affect the reward condition and had no effect on test performance of students (differences not significant at 10% level).

ments. None of the estimated treatment effects are significantly different from one another. Typically, the material cost of non-financial incentives is low – in our case one trophy cost approximately $3. Hence, among this population of students, non-financial incentives are a much more cost effective way of improving student performance than is paying cash.

Table 7 shows the main results from CPS.[22] While financial incentives framed as gains have no impact on performance, there is a weak positive effect for non-financial incentives. Robust positive effects are elicited when framing either financial incentives or non-financial incentives as losses. These interventions raise student performance by $0.1 - 0.13$ standard deviations. The effect sizes of financial and non-financial incentives are roughly equal so that non-financial incentives again turn out to be more cost-effective. All incentives work significantly better than the low financial incentive. There are no statistically significant differences between the other incentive treatments. The pure motivational statement to improve (Control - Statement) does not significantly affect performance.

In order to explore the importance of discount rates, we implemented a delayed version of the four primary 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). To isolate the effect of timing, we pool all the immediate and all the delayed treatments, respectively. Table 8 shows that immediate rewards increase performance on average by $0.07 - 0.08$ standard deviations. Delayed rewards, in comparison, have no effect on performance. The difference between the two treatments is statistically significant at the $p < .05$ level.

Of course, there might be important heterogeneities that determine the effectiveness of the different rewards and that can shed further light into some of the divergent findings between the different settings. Tables 9-11 examine treatment effects by age, test subject,

[22]The control group pools Control - No Statement with students who received a similar treatment ("Control Delayed"). The only difference between the two groups is that students in the Control Delayed condition were told that they would learn their test scores "one month after the test" instead of "immediately after the test" (see scripts in Appendix A). We cannot reject the hypothesis that both groups performed identically at a 10% level.

and gender.[23]

We first examine age. 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 9 estimates treatment effects in CPS for younger students (grades 2-4) and older students (grades 5-8).[24] Younger students respond to all incentives other than the low financial incentive. There are no statistically significant differences between the different treatments and the effect sizes are larger than in the pooled regression. On average, the incentives increase performance by 0.2 standard deviations. Again, the magnitude of these effects is substantial. Older students, in comparison, only respond to financial incentives framed as a loss. Hence, non-financial incentives may be a cost-effective alternative to financial incentives but only with younger children.

We also examine treatment effects by test subject. Previous evidence suggests that when incentives have an effect, the impact on math are higher.[25] Table 10 presents treatment effects separately for reading and math in CPS. In line with Bettinger (2010), we find that incentives have larger effects on math tests. All incentives have a statistically significant positive effect on math test results. There are no statistically significant differences between the effect sizes of the different treatments.[26] The point estimates in math are also much bigger than in the pooled regression and resemble those that we found for Chicago Heights elementary schools where only math was incentivized. Looking at the reading results, on the other hand, only non-financial loss has a positive impact on student performance while financial low even decreases performance.

[23]We also examine heterogenous 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.

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

[25]Bettinger 2010, for example, finds that incentives increase math scores by 0.15 standard deviations, but have no effect on reading, social science and science tests.

[26]As we did not balance treatments with respect to subject, we have no observations for low financial incentives in math.

A further source of heterogeneity might be gender. Evidence on the effect of incentives by gender is mixed with longer term studies tending to find larger effects on girls (e.g. Angrist, Lang, and Oreopoulos 2007, Angrist and Lavy 2007) and shorter term studies finding larger effects among boys, particularly in the context of competition (Gneezy and Rustichini 2004, Gneezy, Niederle, and Rustichini 2001). Such effects might be driven by differences in time preferences between boys and girls.[27] In both Bloom and CPS, treatment effects are larger for boys than for girls (Table 11).[28] This is in line with the literature that shows that boys are more responsive to short-term incentives than girls. Interestingly, there is some evidence that Interestingly, there is some evidence that girls' performance may decrease in response to low level incentives (financial low and control statement). These results are consistent with evidence from other domains that girls may be more intrinsically motivated than boys and thus will be less responsive to high powered incentives and may also experience crowding out effects in response to low powered incentives.

4 Conclusion

In this study, we explore the short-term effects of financial incentives as well as of non-financial rewards on student effort and performance. We vary the sizes of the rewards as well as their framing as either gains or losses. This paper reports findings from 6 waves of a large-scale field experiment in 3 settings that differ in terms of scale, age, and test subject incentivized. To the best of our knowledge this is the first study that addresses short-term incentives in such an encompassing way, not only looking at a large variety of incentives but also testing the robustness of the results (replicability) as well as their generalizability (scaling up from small to large districts, elementary versus high-school).

We find that incentives affect student performance, although there is substantial variation across settings. Low financial incentives ($10) only work in one setting, whereas high financial

[27] Bettinger and Slonin (2007) and Castillo et al. (2011) show that boys are more impatient than girls.
[28] Again, the sample size in Chicago Heights is too small to conduct a similar analysis.

incentives ($20) worked well in 2 out of 3 settings. Financial incentives presented as losses have consistently large and significant effects on performance (though these effects are not statistically significant from incentives in the gain domain).

Overall, incentives work best at Chicago Heights elementary schools. This is a small district with which we have a long-standing research relationship. The results in Chicago Public Schools (CPS) are less strong, which might be due to the greater variation in schools and backgrounds of students in this sample. On the other hand, the divergence in findings between Chicago Heights and CPS elementary schools could be explained by the fact that our research team had a long-term presence in Chicago Heights, but not in CPS. The children in CPS might therefore be less certain that these unusual rewards would actually be handed out. This same explanation might also explain why losses works better than gains at CPS while both work equally well at Bloom in Chicago Heights. In addition to the mechanisms of loss aversion and framing, the loss treatments might have affected behavior as they render the rewards more salient and probably increase students' trust in the likelihood of payout.

Nonfinancial rewards in the form of trophies have a statistically significant positive effect. The effect size varies from 0.25 at Chicago Heights elementary schools to 0.07 at CPS elementary schools. Considering the low financial value of trophies, using non-financial rewards may be more cost-effective for schools than the use of financial incentives. Again there is no statistically significant difference between implementing trophies as gains or losses. Interestingly, trophies are significant even in the gain domain at CPS while high financial rewards are not. Again, this might be a hint that students did not trust that they were actually going to receive $20 because financial rewards are highly uncommon. Trophies and other nonfinancial incentives, in comparison, are frequently used in all schools.

The magnitude of the impact of the incentives on that day's test are quite large: approximately $0.12 - 0.22$ standard deviations, which is similar to effect sizes achieved through a one-standard deviation increase in teacher quality (Rockoff, 2004; Hanushek and Rivkin, 2005; Kane and Staiger, 2008) or 20% reductions in class size (Krueger 1999).

We find that all motivating power vanishes, when we implement the same rewards in a delayed version in which we tell students that they will receive their reward one month after the test, rather than immediately. While this is in line with previous research highlighting the high discount rates of children, it poses a challenge for educators. Typically, test results of the state-wide assessment tests are only available 1-2 months after the administration of the tests.

We find that all motivating power vanishes, when we implement the same rewards in a delayed version where we tell students that they will learn about their test results and, hence, their reward one month, rather than immediately, after the test. While this is in line with previous research highlighting the high discount factor of children, it posesa challenge for educators. Typically, test results of the state wide assessment tests are only available 1-2 months after the test was taken by the children.

Overall, we conclude that both financial and non-financial incentives can serve as useful tools to increase student effort and motivation on otherwise low-stakes assessment tests. Nevertheless, there are important sources of heterogeneity that affect the effectiveness of rewards that need to be taken into account. This analysis is only a first step and says nothing about the long-term consequences of using such rewards. In future work we plan to address potential long-term consequences of rewards and spillover effects over subjects and time.

References

Angrist, Joshua D., Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Columbia." *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): 1-29.

Angrist, Johua D., and Victor Lavy. 2009. "The Effect of High-Stakes High School Achieve-

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

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

Ball, Sheryl, Cathrine C. Eckel, Philip J. Grossman and William Zame. 2001. "Status in Markets." *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, Leight L. Linden and Francisco Perex-Calle. 2008. "Conditional Cash Transfers in Education: Design Features, Peer and Sibling Effects: Evidence from a Randomized Experiment in Columbia." NBER Working Paper No. 13890.

Baumert, Jurgen 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: 441-462.

Behrman, Jere, 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.

Berry, James. 2009. "Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India." Manuscript.

Betrand, Marianne. 2010. "New Perspectives on Gender." Handbook of Labor Economics Volume 4b, Chapter 17: 1546-1592.

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 in Children: Evidence from Experimental Economics." *Journal of Public Economics*, 91(1-2): 343-363.

Blanes i Vidal, Jordi and Mareike Nossol. 2009. "Tournaments Without Prizes: Evidence from Personnel Records ." Working Paper.

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*, forthcoming.

Croson, Rachel and Uri Gneezy. 2009. "Gender Differences in Preferences" *Journal of Economic Literature*, 47(2): 448-474.

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*, 22: 627-668.

Della Vigna, Stefano, John A. List and Ulrike Malmendier. 2011. "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Economics*. Forthcoming.

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

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 Acheivement: Evidence from Randomized Trials." *Quarterly Journal of Eonomics*, forthcoming.

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

Gneezy, Uri and Aldo Rustichini. 2000b. "A Fine is a Price." *Journal of Legal Studies*, 29: 1-18.

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

Gneezy, Uri and Aldo Rustichini. 2004. "Gender and Competition at a Young Age." *The American Economic Review*, XX.

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. 2007. "A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program." http://works.bepress.com/c_kirabo_jackson/1.

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

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.*

Krawacyzyk, 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. Forthcoming. "Incentives to Learn." *The Review of Economics and Statistics.*

Krueger, Alan B. and Diane M. Whitmore. 2001. "The Effect of Attending a Small Class int eh Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *The Economic Journal*, 111(468): 1-28.

Leuven, Edwin, Hesel Oosterbeek and Bas van der Klaauw. Forthcoming. "The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment." *Journal of the European Economic Association.*

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

ONeil, Harolf F., Jr., Sugrue, Brenda, and Baker, Eva L. 1996. "Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance." *Educational Assessment*, 3: 135-157.

ONeil, Harolf F., Jr., Sugrue, Brenda, and Baker, Eva L., 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.

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

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

Tran, Anh and Richard Zeckhauser. 2009. "Rank as an Incentive." Harvard Working Paper rwp09-019.

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

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 (Trophy)				X	X
Financial Delayed				X	
Non-financial Delayed				X	
Financial Loss Delayed				X	
Non-financial Loss Delayed				X	
Number of subjects					

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 $10 treatment are pooled with different, but very similar treatments that add the statement that a student's improvement will be compared to 3 other students with similar past scores (scripts denotes with "comparison" in section A). This comparison statement did not affect the reward condition and had no effect on test performance of students (differences not significant at 10% level).

[b] Similarly, we pool Control-Statement with students who received a similar treatment ("Control Delay"). The only difference between the two groups is that students in the Control Delay condition were told that they would learn their test scores "one month after the test" instead of "immediately after the test" (see scripts in Appendix A). We cannot reject the hypothesis that both groups performed identically at a 10% level.

Table 2: Baseline Characteristics by Treatment Group: Bloom

	Control	Low	High	Loss
N	285	166	324	154
Past Score	.112	.086	-.07	.174
	(.954)	(.9)	(.956)	(1.04)
Female	.523	.524	.435	.468
	(.5)	(.501)	(.497)	(.501)
Black	.586	.452**	.556	.468
	(.493)	(.499)	(.498)	(.501)
Hispanic	.288	.422**	.306	.318
	(.453)	(.495)	(.461)	(.467)
White	.046	.072	.049	.058
	(.209)	(.26)	(.217)	(.235)
Free/Reduced Lunch	.716	.711	.701	.74
	(.452)	(.455)	(.459)	(.44)

Note: The table reports means for each group. Standard devations are displayed in parentheses. Asteriks next to coefficients indicate a difference of means (compared to control) significant at the 10/5/1 percent level. Standard errors are clustered by class.

Table 3: Baseline Characteristics by Treatment Group: Chicago Heights

	Control	Financial Low	Financial High	Nonfinancial
N	179	165	30	69
Prev. test score	-0.511	-0.510	-0.399	-0.682
	(0.057)	(0.061)	(0.195)	(0.093)
Female	0.503	0.497	0.433	0.449
	(0.037)	(0.039)	(0.092)	(0.060)
African-American	0.497	0.461	0.300**	0.290***
	(0.037)	(0.039)	(0.085)	(0.055)
Hispanic	0.391	0.461	0.633**	0.623***
	(0.037)	(0.039)	(0.089)	(0.059)
Other race	0.112	0.079	0.067	0.087
	(0.024)	(0.021)	(0.046)	(0.034)
Free lunch	0.877	0.891	0.900	0.928
	(0.025)	(0.024)	(0.056)	(0.031)
Grade	6.179	5.133***	5.400**	5.072***
	(0.146)	(0.113)	(0.252)	(0.148)
Math	1.000	1.000	1.000	1.000
	(0.000)	(0.000)	(0.000)	(0.000)

Note: The table reports means for each group. Standard errors are displayed in parentheses. Asterisks next to coefficients indicate a difference of means (compared to Control) significant at the 10/5/1 percent level. Performance is standardized with the population of Illinois students who took the ThinkLink tests. Free lunch eligibility is a proxy for family income.

Table 4: Baseline Characteristics by Treatment Group: CPS

	Control No Statement	Statement	Financial Low	Financial High	Non-Financial	Financial Loss	Non-Financial Loss
N	2315	798	106	660	550	748	720
Baseline Scores	-0.183	-0.135	-0.276	-0.117	-0.046*	-0.192	-0.151
	(0.933)	(0.856)	(0.837)	(0.920)	(0.891)	(0.940)	(0.913)
Subject	1.718	1.636	2.000	1.674	1.801	1.720	1.611
	(0.450)	(0.528)	(0.000)	(0.469)	(0.399)	(0.503)	(0.488)
Grade	5.025	5.259	5.623	5.326	5.204	4.785	5.028
	(1.871)	(1.845)	(1.355)	(1.725)	(1.967)	(1.919)	(1.887)
Gender	0.498	0.524	0.692***	0.547**	0.509	0.577***	0.478
	(0.500)	(0.500)	(0.464)	(0.498)	(0.500)	(0.494)	(0.500)
Race	0.923	0.964***	0.972*	0.962***	0.984***	0.959***	0.963***
	(0.267)	(0.187)	(0.167)	(0.191)	(0.127)	(0.199)	(0.190)
Free Lunch	0.926	0.960***	0.981**	0.967***	0.969***	0.956***	0.965***
	(0.262)	(0.196)	(0.137)	(0.180)	(0.173)	(0.205)	(0.183)

Note: The table reports means for each group. Standard deviations are displayed in parentheses. Asterisks next to coefficients indicate a difference of means (compared to Control No Statement) significant at the 10/5/1 percent level. As roughly 95% of the children in CPS are black, *race* codes children as either black or non black. Free lunch eligibility is a proxy for family income.

Table 5: Treatment Effects on Test Score Improvement: Bloom

	I	II	III	IV
Financial Low	-0.047	0.048	0.012	0.003
	(0.070)	(0.133)	(0.146)	(0.154)
Financial High	0.202***	0.170***	0.180***	0.173**
	(0.063)	(0.062)	(0.066)	(0.068)
Financial Loss	0.158**	0.216***	0.188**	0.173**
	(0.059)	(0.051)	(0.071)	(0.069)
Session	Yes	Yes	Yes	Yes
Past score & treatment		Yes	Yes	Yes
Teacher			Yes	Yes
Demographics				Yes
Observations	825	825	825	825
Classes	38	38	38	38

Note: The table reports OLS estimates for pooled sessions in standard deviation units. Robust standard errors clustered by school&grade are reported in parentheses. Asterisks next to coefficients indicate a significance level of 10/5/1 percent. The pooled control group was omitted from the regression. Columns I-V report coefficients on the treatment groups. Column I only contains controls for session. Column II adds controls for past score (score, score squared and score cubed) and past treatment. (score, score squared and score cubed). Column III adds controls for teacher. Column IV includes all previous controls and adds controls for demographics (gender, race/ethnicity and free/reduced lunch status).

Table 6: Treatment Effects on Test Score Improvement: Chicago Heights

	I	II	III	IV
Financial Low	0.205**	0.182**	0.248**	0.241**
	(0.094)	(0.079)	(0.092)	(0.087)
Financial High	0.456***	0.375	0.399**	0.382**
	(0.145)	(0.257)	(0.180)	(0.183)
Nonfinancial	0.317***	0.243**	0.266**	0.255**
	(0.092)	(0.088)	(0.103)	(0.093)
Constant	-0.155**	-0.289***	-0.173	-0.236
	(0.062)	(0.059)	(0.136)	(0.185)
Past score		Yes	Yes	Yes
Grade & school			Yes	Yes
Demographics				Yes
Observations	423	423	423	423
Classes	24	24	24	24

Note: The table reports OLS estimates for pooled treatments in standard deviation units. Robust standard errors clustered by school & gradel are reported in parentheses. Asterisks next to coefficients indicate a significance level of 10/5/1 percent. The pooled control group was omitted from the regression. Columns I-V report coefficients on the treatment groups. Column I contains no controls. Column II contains controls for standardized scores of the previous test (score, score squared and score cubed). Column III adds controls for grades and schools. Column IV includes all previous controls and adds adds controls for demographics (gender, race/ethnicity and free/reduced lunch status).

Table 7: Treatment Effects on Test Score Improvement: CPS

	I	II	III	IV
Control Statement	-0.001	0.016	0.021	0.012
	(0.041)	(0.042)	(0.044)	(0.043)
Financial Low	-0.080	-0.032	-0.076	-0.073
	(0.071)	(0.058)	(0.064)	(0.061)
Financial High	0.058	0.085	0.067	0.064
	(0.054)	(0.054)	(0.051)	(0.048)
Non-Financial	0.026	0.064	0.069	0.071*
	(0.049)	(0.045)	(0.042)	(0.040)
Financial Loss	0.120**	0.135***	0.127***	0.128***
	(0.046)	(0.041)	(0.043)	(0.042)
Non-Financial Loss	0.103*	0.129**	0.132***	0.129***
	(0.052)	(0.051)	(0.046)	(0.046)
Session	Yes	Yes	Yes	Yes
Past Score & treatment		Yes	Yes	Yes
School, subject & grade			Yes	Yes
Demographics				Yes
Students	5439	5439	5439	5434
School-grades	167	167	167	167

Note: The table reports OLS estimates for pooled treatments in standard deviation units. Robust standard errors clustered by school &vgrade are reported in parentheses. Asterisks next to coefficients indicate a significance level of 10/5/1 percent. The "no statement" control group was omitted from the regression. Columns I-V report coefficients on the treatment groups. Column I only controls for session. Column II contains controls for standardized scores of the previous test (score, score squared and score cubed) and past treatment. Column III adds controls for school, grade and subject (math or reading). Column IV includes all previous controls and adds adds controls for demographics (gender, race/ethnicity and free/reduced lunch status).

Table 8: Effects of Reward Timing (CPS)

	I	II	III	IV
Immediate	0.065*	0.077**	0.081**	0.079**
	(0.034)	(0.033)	(0.033)	(0.032)
Delayed	-0.056	-0.060	-0.034	-0.036
	(0.084)	(0.075)	(0.076)	(0.074)
Session	Yes	Yes	Yes	Yes
Past Score & treatment		Yes	Yes	Yes
School, subject & grade			Yes	Yes
Demographics				Yes
Students	6027	6027	6027	6015
School-grades	169	169	169	169

Note: The table reports OLS estimates for pooled treatments in standard deviation units. Robust standard errors clustered by school & grade are reported in parentheses. Asterisks next to coefficients indicate a significance level of 10/5/1 percent. Columns I-V report coefficients on the treatment groups. The "no statement" control group was omitted from the regression. Columns I-V report coefficients on the treatment groups. Column I only controls for session. Column II contains controls for standardized scores of the previous test (score, score squared and score cubed) and past treatment. Column III adds controls for school, grade and subject (math or reading). Column IV includes all previous controls and adds adds controls for demographics (gender, race/ethnicity and free/reduced lunch status).

Table 9: Treatment Effects By Grade (CPS)

	2nd - 4th graders		5th - 8th graders	
	(1)	(2)	(3)	(4)
Financial Low	-0.027	-0.168	-0.083	-0.034
	(0.074)	(0.133)	(0.055)	(0.076)
Financial High	0.129*	0.189**	0.008	0.014
	(0.074)	(0.075)	(0.069)	(0.057)
Non-Financial	0.113	0.197***	-0.066	-0.032
	(0.075)	(0.069)	(0.053)	(0.047)
Financial Loss	0.115	0.235***	0.133**	0.105**
	(0.078)	(0.084)	(0.052)	(0.043)
Non-Financial Loss	0.185***	0.218***	0.033	0.093
	(0.062)	(0.057)	(0.076)	(0.060)
Session	Yes	Yes	Yes	Yes
Other covariates		Yes		Yes
Students	2267	2264	3172	3170
School-grades	81	81	99	99

Note: The table reports OLS estimates for pooled sessions in standard deviation units. Robust standard errors clustered by school & grade are reported in parentheses. Asterisks next to coefficients indicate a significance level of 10/5/1 percent. The pooled control group was omitted from the regression. Columns (1) and (3) control for session. Columns (2) and (4) adds controls for past score (score, score squared and score cubed) and past treatment, controls for teacher, and for demographics (gender, race/ethnicity and free/reduced lunch status).

Table 10: Treatment Effects by Subject (CPS)

	Math		Reading	
	(1)	(2)	(3)	(4)
Control Statement	0.060	0.166*	-0.029	-0.036
	(0.075)	(0.083)	(0.047)	(0.045)
Financial Low	—	—	-0.105	-0.137**
	—	—	(0.075)	(0.062)
Financial High	0.124	0.190*	0.026	0.031
	(0.112)	(0.102)	(0.059)	(0.045))
Non-Financial	0.161	0.352***	-0.011	0.026
	(0.109)	(0.127)	(0.053)	(0.038)
Financial Loss	0.262***	0.329***	0.057	0.065
	(0.076)	(0.084)	(0.055)	(0.049)
Non-Financial Loss	0.178**	0.222***	0.056	0.112**
	(0.082)	(0.079)	(0.061)	(0.052)
Session	Yes	Yes	Yes	Yes
Other covariates		Yes		Yes
Students	1656	1652	3752	3751
School-grades	65	65	119	119

Note: The table reports OLS estimates for pooled treatments in standard deviation units for math and reading tests. Robust standard errors clustered by school & grade are reported in parentheses. Asterisks next to coefficients indicate a significance level of 10/5/1 percent. Columns (1) and (3) control for session. Columns (2) and (4) add controls for past score (score, score squared and score cubed) and past treatment, controls for teacher, and for demographics (gender, race/ethnicity and free/reduced lunch status).

Table 11: Treatment Effects by Gender (Bloom & CPS)

	Female		Male	
	(1)	(2)	(3)	(4)
Panel A: Bloom				
Financial Low	-0.207*	-0.229	0.135	0.270
	(0.104)	(0.145)	(0.102)	(0.171)
Financial High	0.182**	0.140	0.220***	0.214**
	(0.082)	(0.086)	(0.077)	(0.069)
Financial Loss	0.082	0.076	0.225**	0.271**
	(0.084)	(0.098)	(0.086)	(0.101)
Session	Yes	Yes	Yes	Yes
Other covariates		Yes		Yes
Observations	412	412	413	413
Classes	38	38	38	38
Panel B: CPS				
Control Statement	-0.027	-0.018	0.025	0.050
	(0.043)	(0.046)	(0.052)	(0.049)
Financial Low	-0.144*	-0.101	—	—
	(0.054)	(0.048)	—	—
Financial High	0.006	0.041	0.114*	0.088
	(0.073)	(0.096)	(0.067)	(0.059)
Non-Financial	-0.016	0.032	0.068	0.114**
	(0.060)	(0.049)	(0.050)	(0.047)
Financial Loss	0.077	0.086*	0.170***	0.175***
	(0.056)	(0.050)	(0.051)	(0.050)
Non-Financial Loss	0.031	0.062	0.171**	0.205***
	(0.047)	(0.043)	(0.071)	(0.064)
Session	Yes	Yes	Yes	Yes
Other covariates		Yes		Yes
Observations	2818	2818	2616	2616
Classes	167	167	166	166

Note: The table reports OLS estimates for pooled sessions in standard deviation units. Robust standard errors clustered by class (Bloom) or school & grade (CPS) are reported in parentheses. Asterisks next to coefficients indicate a significance level of 10/5/1 percent. The pooled control group was omitted from the regression. Columns (1) and (3) control for session. Columns (2) and (4) adds controls for past score (score, score squared and score cubed) and past treatment, controls for teacher, and for demographics (gender, race/ethnicity and free/reduced lunch status).

A Appendix: Administrator Scripts

A.1 Experiments 1 & 2: Bloom High School, Chicago Heights

Bloom 2009

For all treatments

To the Proctor:

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):

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 2009

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 High - $20 - Loss 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 Experiment 3: Primary Schools, Chicago Heights

Control

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 - 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 - Immediate

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 - Immediate

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.

Nonfinancial - Trophy - Immediate

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 Experiments 4 & 5: Chicago Public School District

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 - Motivational 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 - Motivational Statement - Delay

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 - Immediate

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 - Immediate

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.

Nonfinancial - Trophy - Immediate

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.

Financial high - $20 - Delay

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.

Nonfinancial - Trophy - Delay

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.

Financial high - $20 - Immediate - Loss

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.

Nonfinancial - Trophy - Immediate - Loss

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.

Financial high - $ 20 - Delay - Loss

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.

Nonfinancial - Trophy - Delay - Loss

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.