American Economic Journal: Economic Policy 2016, 8(4): 183–219
http://dx.doi.org/10.1257/pol.20130358

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

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

We explore the power of behavioral economics to influence the level of effort exerted by students in a low stakes testing environment. We find a substantial impact on test scores from incentives when the rewards are delivered immediately. There is suggestive evidence that rewards framed as losses outperform those framed as gains. Nonfinancial incentives can be considerably more cost-effective than financial incentives for younger students, but are less effective with older students. All motivating power of incentives vanishes when rewards are handed out with a delay. Our results suggest that the current set of incentives may lead to underinvestment. (*JEL* D03, H75, I21, I28)

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 relatively limited inroads, however, is education. This is puzzling since it is an area where the insights gained from behavioral economics might be especially great (Lavecchia, Liu, and Oreopoulos 2014). In this study, we use a series of field experiments to explore whether interventions informed by behavioral economics lead students to exert more effort on a low stakes test, and if so, what broader implications these results have for education policy.

*Levitt: Saieh Hall for Economics, University of Chicago, Room SHFE 434, 1126 E 59th Street, Chicago, IL 60637 and NBER (e-mail: slevitt@uchicago.edu); List: Saieh Hall for Economics, University of Chicago, Room SHFE 415, 1126 E 59th Street, Chicago, IL 60637 and NBER (e-mail: jlist@uchicago.edu); Neckermann: Saieh Hall for Economics, University of Chicago, SHFE 358, 1126 E 59th Street, Chicago, IL 60637 and ZEW (e-mail: sneckermann@uchicago.edu); Sadoff: Rady School of Management, University of California San Diego, Wells Fargo Hall, Room 4W121, 9500 Gilman Drive 0553, La Jolla, CA 92093 (e-mail: ssadoff@ucsd.edu). 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, Jeannine van Reeken, 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.

†Go to http://dx.doi.org/10.1257/pol.20130358 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

Our contribution is two-fold. First, we demonstrate that behavioral economics can help shed light on our understanding of the education production function and perhaps the design of educational interventions. Second, we demonstrate a model for using "basic research" as a way to inform policymaking. We do this by developing an experimental design that allows us to identify and explore a single input—effort—of the education production function. We then conduct a series of experiments that begin with proof-of-concept and gradually scale up to test generalizability across different settings, grades, subjects, and student characteristics. Our work is not itself a ready-made program, but can potentially inform a wide range of interventions. We argue there should be a larger role for this kind of research in the policymaker's toolkit.

One of the biggest puzzles in education is why investment among many students is so low given the high returns. One explanation is that the current set of long-run returns does not sufficiently motivate some students to invest effort in school. If underinvestment is a problem, then there is a role for public policy in stimulating investment. Towards that end, a number of papers in recent years have examined the effects of monetary rewards on a variety of measures including school enrollment, attendance, behavior, grades, test performance, and matriculation. Examples include Progresa in Mexico, which 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 et al. 2011). Other programs have based rewards on overall school performance (see Angrist, Bettinger, and Kremer 2006; Levitt, List, and Sadoff 2010; Leuven, Oosterbeek, and van der Klaauw 2010; Fryer 2011; Patel et al. 2013). Results have varied across settings, but overall these financial incentives have been associated with a modest improvement in educational outcomes.[1]

Although the incentive structure and performance measures of previous programs have varied, they tend to share the following features. First, they offer rewards as *gains*. That is, students can only receive and experience the reward after exerting effort and meeting the performance criteria. Second, they primarily employ monetary rewards. Third, the incentives are typically announced well in advance of the incentivized task with a delay of weeks or months between the time students must exert effort and the time they receive rewards.

In this paper, we extend that line of research by focusing explicitly on one dimension of the production function (effort exerted while taking an exam), and by drawing on three areas of behavioral economics to try to improve the cost-effectiveness of these interventions: loss aversion, nonmonetary rewards, and hyperbolic discounting.[2] A key feature of the education investment function is that in order to

[1] In the settings most similar to ours, Bettinger (2012) finds that incentives of up to $20 have a significant impact on third through sixth graders' performance in math but no impact on reading, social science, or science. Fryer (2011), in comparison, finds no effect on either math or reading test scores of offering incentives of up to $30 to fourth graders and $60 to seventh graders.

[2] Previous work drawing on behavioral economics in education has primarily explored the role of information. These studies have found that increased information and understanding can improve decision-making and outcomes in educational attainment (Jensen 2010), school achievement (Nguyen 2008; Bergman 2012), school choice (Hastings and Weinstein 2008), college enrollment (Dynarski and Scott-Clayton 2008; Bettinger 2012; Hoxby and Turner 2013), and financial planning (Hastings and Mitchell 2011).

experience the long-run returns to schooling, students must make sustained investments in human capital that require exerting effort on tasks that often have relatively low returns in the near term, such as paying attention in class, completing a daily assignment, or focusing on a practice test. While these low stakes effort decisions are among the primary investment decisions students make in education, they are not well understood. Effort is usually difficult to measure directly. And policies aimed at increasing achievement often cannot disentangle the effect of an intervention on student motivation and effort from its effect on learning and human capital accumulation.

This is particularly important when we find that a policy has little or no effect. Is it because the intervention, if wholeheartedly adopted by students, does not promote increased learning, or is it because students did not invest effort into the program? For example, in his study of incentive interventions in multiple US school districts, Fryer (2011) attributes his largely null findings in part to students' lack of understanding of the production function. That is, even if students are motivated by the incentives they do not know how to respond productively to them. An alternate explanation that the experimental design cannot rule out is that students are simply not motivated sufficiently by the incentives to invest effort into improving performance. We set out to understand what motivates students to exert effort, which is the first necessary condition for building human capital. To do this, we incorporate insights from behavioral economics into the standard economic framework.

Our study evolved in several steps. In the first wave of our field experiment, we wanted proof-of-concept that rewards offered immediately before and delivered immediately after an incentivized task could motivate students to exert greater effort. Typically, rewards are offered at the end of the term or year and at the earliest on a monthly basis. Numerous studies 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). One cause of high discount rates is hyperbolic time preferences, overweighting the present so much that future rewards are largely ignored (e.g., Strotz 1955; Laibson 1997). Such preferences can lead to underinvestment when (as in education) the returns to achievement are largely delayed.[3] If students are sufficiently myopic, they will respond more strongly to rewards with very short time horizons (e.g., minutes) compared to incentives extending over several months or years.

In order to test this, we needed a setting in which increased effort would move a performance measure; and the measure needed to be available immediately after students exerted effort. We therefore chose to offer incentives on a low stakes computer-based diagnostic test in which results were available immediately after students completed the test. In order to ensure that we were identifying motivation and

[3] Previous studies find a negative correlation between hyperbolic discount rates and educational outcomes (Kirby et al. 2002; Kirby, Winston, and Santiesteban 2005; Castillo et al. 2011). Similarly, Mischel, Shoda, and Rodriguez (1989) find that measures of ability to delay gratification in early childhood are predictive of longer-term academic achievement. Cadena and Keys (2015) and Oreopoulos (2007) find evidence that impatience may partially explain school dropout behavior.

effort, we announced incentives directly before the incentivized task, so the only channel of improvement is through increased effort and focus during the exam.[4]

In addition, we vary the size of the reward in order to distinguish students' ability to improve performance from their motivation to do so. If students require sufficient motivation to exert effort (e.g., because effort costs are high), they may respond to high-powered incentives but not to low-powered incentives. On the other hand, if baseline effort is high (i.e., students are close to their effort frontier), or if students do not understand the production function (i.e., what types of effort will improve performance) they may be unable to respond to incentives regardless of their motivating power.

In our second and third waves, we explored the design of incentives within our basic framework of immediate rewards. Among older students in our original school district, we designed rewards framed as losses rather than gains. Among younger students in a second school district, we introduced nonfinancial rewards.

With respect to loss aversion, a large literature demonstrates that some individuals have reference dependent preferences wherein they respond more strongly to losses than gains. 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 2003), are broadly consistent with a notion of *loss aversion* from Kahneman and Tversky's (1979) prospect theory. If this is true for students, then framing incentives as losses rather than gains may increase the impact of the intervention. While similar framing mechanisms have been widely explored in the lab, there are to-date only a few studies that experimentally test loss aversion in the field.[5] Because these types of rewards are novel in schools, we tested them first among older students to ensure that they were logistically feasible.

With respect to nonfinancial rewards, we build on a growing area of research demonstrating their motivational power (e.g., Bradler et al. forthcoming; Frey 2007; Kosfeld and Neckermann 2011; Ashraf, Bandiera, and Jack 2014; Jalava, Joensen, and Pellas 2014). Nonfinancial rewards potentially operate through a range of mechanisms including status, self-image concerns, and relative performance feedback that have been shown to affect behavior.[6] These types of non-pecuniary benefits

[4] To the best of our knowledge, a study produced concurrently to ours—Braun, Kirsch, and Yamamoto (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 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. 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, Sugrue and Baker (1995); O'Neil et al. (2005) 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).

[5] Previous field experiments have, for example, tested the effect of the loss frame in marketing messages on product demand (Ganzach and Karsahi 1995; Bertrand et al. 2010). In the context of incentives, Hossain and List (2012) find that framing bonuses as losses improves the productivity of teams in a Chinese factory. In studies run concurrently to ours, Fryer, Jr. et al. (2012) find that framing bonuses as losses can improve teacher performance, while List and Samek (2015) find no framing effects for student incentives to make healthy food choices. Volpp et al. (2008) find that deposit contracts in the loss domain improve weight loss compared to an unincentivized control group (the study does not include incentives in the gain domain).

[6] See, among others, Ball et al. (2001) and Huberman, Loch, and Önçüler (2004) on status; Blanes i Vidal and Nossol (2011), Tran and Zeckhauser (2012), and Barankay (2011) on relative performance feedback; and Ariely,

could be especially potent in the educational context. The implication of this line of research is that, in contrast to standard models, some students may be willing to exert more effort for a trophy worth \$3 than they are for \$3 in cash. Non-pecuniary incentives are also attractive because they are already commonly used in schools, which tend to be more comfortable rewarding students with trophies, certificates, and prizes than they are with using cash rewards. Despite their widespread prevalence, however, the effectiveness of nonfinancial incentives is largely untested—particularly in terms of cost-effectiveness relative to monetary rewards.[7] We introduced these rewards among younger children because they may be particularly responsive to nonfinancial incentives as they are often more familiar with them than they are with financial rewards.

After confirming that the incentive designs were feasible and effective, we scaled them up in the fourth and fifth waves in a third school district to test for generalizability and to explore heterogeneous effects with a larger sample. The larger sample size also allowed us to compare the effects of immediate incentives to identical rewards offered with a short delay (of one month). We implemented the delayed variant for both theoretical and policy-related reasons. First, it was important to confirm that delaying incentives reduces their effectiveness, as we had hypothesized in the motivation of our design. Second, schools were interested in testing the delay because on some tasks it is logistically difficult for them to distribute rewards immediately. For example, the results of state standardized tests are generally not available until several weeks or months after students take the exam.

Altogether, we test our incentive designs in a field experiment involving over 5,700 elementary, middle, 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.[8]

We find that large incentives delivered immediately, whether financial or nonfinancial, have a significant impact on test performance of about a tenth of a standard deviation. In stark contrast, rewards delivered with a one month delay have no impact, nor do small financial rewards. Indeed, there is suggestive evidence that small financial rewards not only have no positive effect on the incentivized test, but also induce negative spillovers on other tests. We find some evidence that framing the interventions as losses rather than gains magnifies their effectiveness.

The design also allows us to uncover some of the underlying heterogeneities that drive the overall effectiveness of reward schemes: younger children are more

Bracha, and Meier (2009) and DellaVigna, List, and Malmendier (2012) on image motivation and social pressure. For individuals who 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).

[7] Exceptions are O'Neil, Sugrue, and Baker (1995) and Baumert and Demmrich (2001), which test both financial and nonfinancial incentives (instructions, feedback, grades) for test performance.

[8] In a similar vein, Braun, Kirsch, and Yamamoto (2011) test a single performance pay incentive among 2,600 students in 59 schools and 7 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).

responsive to nonfinancial rewards than older children; effects are somewhat stronger among boys than girls; and overall, the incentives work better on math than on 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 potentially have 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. Relatively lower baseline effort among certain groups of students can create important biases in measures of student ability, teacher value added, school quality, and achievement gaps.[9] Understanding the extent to which performance gaps are due to lower effort rather than lower ability is crucial for the design of effective educational interventions: the former requires an intervention that increases student motivation, the latter requires an intervention that improves student knowledge and skills.

In addition, the diagnostic tests in our experiments are similar in nature to many of the low stakes tasks students must engage in daily in order to accumulate human capital. If delays in rewards reduce student effort in our context, it would seem likely that the typical 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) induces suboptimal effort in general. This study provides insights into which instruments may be fruitful in stimulating student effort more broadly.

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

I. Experimental Design and Implementation

The field experiment was carried out in five waves in three low-performing school districts in and around Chicago: Bloom Township (Bloom), Chicago Heights (CH), and Chicago Public Schools (CPS). We incentivized low stakes tests that students do not generally prepare for or have any external reason to do well on. They are computer-based and last between 15–60 minutes with students' results available immediately after the test ends.[10]

[9] Baumert and Demmrich (2001) and Braun, Kirsch, and Yamamoto (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 Chicago Public Schools (CPS) 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. Attali, Neeman, and Schlosser (2011), however, find that the performance of white students falls more than students of other races when moving from a high stakes to a low stakes environment.

[10] The tests are designed to be aligned with the high stakes state standardized test that students take in their respective school district and grade. Students in the same school district, grade, and testing period take the same test. Students are not time-constrained on any of the tests. In fact many of the teachers and principals noted that testing took much longer than usual when students were offered an incentive. Unfortunately we do not have access to measures of how long students spent on the test.

In sessions where we offered rewards, immediately before testing began, the test administrator announced the incentive and told students that they would receive the reward immediately (or in some treatments a month) after the test ended if they improved upon their score from a prior testing session.[11] Immediately after the test ended, we handed out rewards privately to qualifying students, except in the case of delayed rewards, which were distributed a month after 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). This allows us to test whether there are effects due to the presence of the experimenters or of merely requesting that the students improve (we did not attend "No Statement" treatments).

As discussed above, the differences in which treatments were tested in the various waves is due to differences in: student age (e.g., we introduced nonfinancial incentives in Chicago Heights elementary schools rather than the Bloom high school under the hypothesis that younger students would be more responsive than older students to the trophies we used); logistical constraints (e.g., we demonstrated the feasibility of incentives framed as gains before introducing incentives framed as losses); district size (e.g., we were able to add the delayed variant of the incentives in CPS); and, our evolving understanding of the incentives' effectiveness (e.g., the final wave includes only the incentives found to be effective in prior waves). The various waves included additional incentive treatments not discussed here. To keep the analysis tractable, this paper reports the results from those incentives that are common across the settings. Information on the additional treatments and their results are available upon request. Scripts for the different treatments can be found in Appendix A. An overview of the treatments conducted is presented in Tables 1 and 2. Below we discuss the details of implementation in each school district.

A. *Bloom*

We ran the first wave of the study in Bloom Township (Bloom), a small school district south of Chicago with approximately 3,000 students. The first wave was conducted in winter and spring 2009 among high school sophomores at one high school in Bloom. The second wave took place in spring 2010 with a new cohort of Bloom sophomores. The experiment took place during regularly scheduled sessions of the STAR Reading Assessment, a low stakes diagnostic test, which is adaptive and lasts about 15 minutes.[12] Students take the tests three times a year in the fall, winter, and spring.

Students received no notice of the incentives prior to the testing sessions. 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 compensation for their participation. We

[11] The researchers gave the test administrator the relevant treatment script before the test session and asked her to read it after giving students the standard testing instructions and just before students began the test.

[12] The correlation between the STAR Reading test and the ACT PLAN (a preliminary ACT administered to tenth graders) is 0.53, significant at the $p < 0.05$ level (Renaissance Learning 2015).

Table 1—Overview of the Experiment

	Bloom	Chicago Heights	Chicago Public Schools (CPS)
Sample	666 tenth grade students (828 observations) in 1 high school randomized at the class level	343 third–eighth grade students in 7 elementary/middle schools randomized at the school-grade level	4,790 second–eighth graders (6,060 observations) in 26 elementary/middle schools randomized at the school-grade level
Period	Winter and spring 2009 (same cohort, wave 1) Spring 2010 (new cohort, wave 2)	Spring 2010 (wave 3)	Fall 2010 (wave 4) and winter 2011 (same cohort, wave 5)
Subject—assessment	Reading—STAR Reading Assessment	Math—ThinkLink Predictive Assessment Series	Math or Reading—Scantron Performance Series
Reward structure	Students receive the reward if they improve upon their fall baseline STAR score.	Students receive the reward if they improve upon their winter baseline ThinkLink score.	Students receive the reward if they improve upon their baseline Scantron score. Spring 2010 serves as the baseline for fall 2010 testing. Fall 2010 serves as the baseline for winter 2011 testing.
Reward timing	Rewards announced immediately before testing by test administrator. Rewards distributed immediately after testing ends.	Rewards announced immediately before testing by test administrator. Rewards distributed immediately after testing ends.	Rewards announced immediately before testing by test administrator. Rewards distributed either immediately after testing ends or one month after testing ends in delayed incentive treatments.

did not specify the incentives and we sent the same consent form to the treatment and control groups. 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 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.

Incentivized students were offered a reward for improving upon their fall baseline score (in the 2009 wave, fall 2008 served as the baseline; in the 2010 wave, fall 2009 served as the baseline). In the first wave, students were offered either a low financial incentive ($10 cash) or a high financial incentive ($20 cash). As we discussed above, the purpose of the first wave was to establish that immediate rewards could motivate greater effort. We varied the size of the reward in order to establish that a high enough incentive could be effective, and to examine students' incentive sensitivity. This helps inform both our understanding of the education production function and how to cost-effectively design incentives. As we discuss below, we found that among high school students the $20 incentive was effective but the $10 incentive was not.

In the second wave, we therefore included only the high financial incentive and compared framing the reward as a gain (as we had tested previously) to framing the reward as a loss. In the gain condition, the test administrator held up the reward ($20 cash) at the front of the room. In the loss condition students received $20 in cash at the start of the testing session, signed a form confirming receipt of the money

Table 2—Overview of the Treatments

		Bloom		Chicago Heights	Chicago Public Schools (CPS)	
		2009	2010	2010	2010	2011
Control treatments:						
Control—no statement	Experimenters are not present during testing and the test administrator makes no additional statements.	X				X
Control—statement	Experimenters are present during testing and the test administrator encourages students to improve on the test.		X	X[a]	X[b]	X
Rewards distributed immediately after testing:						
Financial low	\$10 cash	X		X	X	
Financial high	\$20 cash	X	X	X	X	X
Nonfinancial	Trophy (cost ~ \$3)			X	X	X
Financial loss	\$20 cash Reward is given to students before testing. Students must return the reward immediately after testing if they do not improve.		X		X	X
Nonfinancial loss	Trophy (cost ~ \$3) Reward is given to students before testing. Students must return the reward immediately after testing if they do not improve.				X	X
Rewards distributed one month after testing:						
Delayed financial high	\$20 cash				X	
Delayed nonfinancial	Trophy (cost ~ \$3)				X	
Delayed financial loss	\$20 cash Reward is given to students before testing. Rewards are collected immediately after testing and redistributed to qualifying students a month after testing.				X	
Delayed nonfinancial loss	Trophy (cost ~ \$3) Reward is given to students before testing. Rewards are collected immediately after testing and redistributed to qualifying students a month after testing.				X	

[a] Control—Statement is pooled with Control—Statement Comparison, which adds 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 percent 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 percent level.

and kept the reward at their computer during testing. They were informed that they would keep the reward if they improved and that they would lose the reward if they did not improve.

Immediately after testing ended, we privately informed students whether they had improved and distributed the cash incentives. In the loss treatment, we collected the upfront rewards from all students at the end of testing and then privately returned rewards to qualifying students. 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). In results that pool the Bloom waves, we pool the Control—No Statement (2009 wave) and Control—Statement (2010 wave) groups. The results are similar across waves (Table 6) and pooling does not affect the results (Appendix B Table B1).

We randomized at the level of English class (which is how the school organized testing), blocking on average class baseline reading score. If the baseline score was not available, we blocked classes by their track: regular, remedial, or honors. In the Bloom 2009 wave, students participated in two testing sessions (winter 2009 and spring 2009), which were each randomized. Thus, some students received the same treatment in both sessions, while others received a different treatment in the two sessions. In cases where students had received incentives in a previous session, there was no reason for them to expect the experiment 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. We examine spillovers to future testing and also present the results by session in order to address this concern. As discussed below in the results section, we find limited evidence that incentive treatments affect subsequent test performance (Table 11). We also find that the results are largely consistent across sessions (Appendix B Table B1).

Table 3 reports summary statistics by treatment group for pretreatment characteristics in Bloom pooling the 2009 and 2010 waves. The pretreatment characteristics include standardized baseline reading score and the following demographics: gender, race/ethnicity, and free or reduced price lunch status, which serves as a proxy for family income. We standardize test scores within session to have mean zero and standard deviation one using the full population of Bloom students. We report tests of differences between individual incentive groups and the control group, as well as tests of equality of means across all groups (in the final column), with standard errors clustered by class. The only significant differences between the control and individual incentive groups are the percentage of black and Hispanic students in the financial low ($10) treatment and the percentage of black students in the financial loss ($20) treatment. There is also imbalance with respect to the overall distribution of black students. As shown in Section II, the results are robust to including controls for pretreatment characteristics.

B. *Chicago Heights*

Like Bloom, Chicago Heights is a small school district south of Chicago with approximately 3,000 students (Chicago Heights elementary and middle schools

Table 3—Baseline Characteristics by Treatment Group: Bloom

	Control	Financial low	Financial high	Financial loss	p-value
Observations	315	177	297	128	
Baseline test score	0.125	0.106	−0.077	0.289	0.567
	(0.966)	(0.903)	(0.972)	(1.035)	
Female	0.506	0.508	0.489	0.487	0.990
	(0.500)	(0.500)	(0.500)	(0.500)	
Black	0.633	0.463	0.588	0.454	0.033
	(0.482)	(0.499)	(0.492)	(0.498)	
Hispanic	0.263	0.401	0.313	0.359	0.137
	(0.440)	(0.490)	(0.464)	(0.480)	
Free or reduced price lunch	0.752	0.734	0.710	0.758	0.907
	(0.432)	(0.442)	(0.454)	(0.428)	

Notes: The table reports group means pooling the Bloom 2009 and Bloom 2010 waves. Standard deviations are reported in parentheses. Baseline score is standardized within testing period to have mean zero and standard deviation one using the full sample of Bloom students. The reported p-value is the probability from a joint F-test that the means are equal to one another, clustering by class.

feed into Bloom High School). The third wave of our study took place in spring 2010 among third–eighth graders in seven schools in Chicago Heights. The experiment took place during the math portion of the ThinkLink Predictive Assessment Series, which is aligned with the state standardized test and lasts about 30 minutes.[13] Students take the test four times per year at the beginning of the school year, and then in the fall, winter, and spring.

As in Bloom, students received no notice of the incentives prior to the testing session. The consent procedures were identical to those described above except that the consent form indicated that students could receive either financial or nonfinancial compensation for their participation. As in Bloom, parents only needed to sign the consent form if they did *not* want their child to participate in the study. Less than 1 percent of parents opted out by returning the form and all eligible students signed the assent form to participate.

Incentivized students were offered one of the following rewards for improving upon their winter 2010 baseline score: financial low ($10 cash), financial high ($20 cash), or nonfinancial (trophy). As discussed above, we introduced nonfinancial rewards among younger students under the hypothesis that they would be more responsive to them than high school students. We tested both low and high financial rewards in order to examine whether younger students were less sensitive than older students to the size of the reward. This also allows us to price out the cost-effectiveness of nonfinancial incentives relative to cash rewards.

In all treatments, the test administrator held up the reward at the front of the room before testing. Immediately after testing, we privately informed students whether

[13] For third–eighth grades, the correlation at the grade level between the ThinkLink assessment and the Illinois Standards Achievement Test (ISAT) is 0.57–0.85 with all correlations significant at the $p < 0.01$ level (Discovery Education 2008).

Table 4—Baseline Characteristics by Treatment Group: Chicago Heights

	Control	Financial low	Financial high	Nonfinancial	p-value
Observations	194	68	29	72	
Baseline test score	−0.551	−0.563	−0.421	−0.682	0.097
	(0.688)	(0.827)	(1.078)	(0.775)	
Grade	6.448	4.279	5.414	5.028	0.325
	(1.949)	(1.195)	(1.402)	(1.222)	
Female	0.448	0.485	0.448	0.458	0.994
	(0.497)	(0.500)	(0.497)	(0.498)	
Black	0.456	0.294	0.310	0.278	0.045
	(0.498)	(0.456)	(0.462)	(0.448)	
Hispanic	0.424	0.603	0.621	0.639	0.006
	(0.494)	(0.489)	(0.485)	(0.480)	
Free or reduced price lunch	0.912	0.897	0.897	0.931	0.471
	(0.283)	(0.304)	(0.304)	(0.253)	
Individualized Education Plan (IEP)	0.108	0.149	0.034	0.097	0.240
	(0.310)	(0.356)	(0.181)	(0.296)	

Notes: The table reports group means. Standard deviations are reported in parentheses. Baseline score is standardized within grade to have mean zero and standard deviation one using the full sample of Illinois students. The reported p-value is the probability from a joint F-test that the means are equal to one another, clustering by school-grade.

they had improved and distributed the incentives. In the nonfinancial treatment we additionally took a photo of qualifying students to be posted in their school. In the control groups, the test administrator encouraged students to improve on the test but did not offer any incentive to do so (Control—Statement). A second control treatment (Control—Statement Comparison) added a statement that we would compare a student's improvement to three other students with similar past scores, with no financial incentive tied to the comparison. In the results below, we pool the two control groups. The comparison statement did not affect test performance at the 10 percent significance level and the results are robust to excluding the comparison treatment from the control group (Appendix B Table B3).

We randomized at the level of school-grade and blocked the randomization on average school-grade baseline math and reading scores, school, grade, and race/ethnicity. Table 4 reports summary statistics by treatment group for pretreatment characteristics. The pretreatment characteristics include standardized baseline math score and the following demographics: grade, gender, race/ethnicity, free or reduced price lunch status, and eligibility for an Individualized Education Plan (IEP), which provides additional services to struggling students.[14] We standardize test scores within grade to have mean zero and standard deviation one using the full population of Illinois students, and cluster standard errors by school-grade. The only significant differences between individual incentive groups and the control group are the proportion of Hispanic students in the nonfinancial treatment. There is also overall imbalance in baseline test scores and the distribution of black and Hispanic students

[14] IEP status was not available for Bloom students and so is not included as a covariate in that setting.

across treatments. As shown in Section II, the results are robust to including controls for pretreatment characteristics.

C. *Chicago Public Schools* (*CPS*)

The final two waves scaled up the Bloom and Chicago Heights experiments and were conducted among second through eighth graders in 26 Chicago Public Schools in fall 2010 and winter 2011. Chicago Public Schools (CPS) is the third largest school district in the United States with approximately 400,000 students. Like Bloom and Chicago Heights, the schools where we ran the experiment are made up of largely low-income, minority students. In CPS, the experiment took place during either the math or reading portion of the Scantron Performance Series, which is a computer-adaptive diagnostic test that is aligned with the state standardized test and lasts about 60 minutes per subject.[15]

As in Bloom and Chicago Heights, students received no notice of the incentives prior to the testing sessions. The consent procedures were identical to those described above except that in CPS, parents needed to sign the consent form in order for their child to participate. Sixty-eight percent of parents returned the signed consent form and, as in previous waves, all students opted into the study by signing the assent form. The analysis only includes students who met the consent criteria. Students who did not meet the consent criteria participated in testing but were not eligible to receive rewards.

Incentivized students were offered a reward for improving their baseline score from the prior testing session (in fall 2010, spring 2010 served as the baseline; in winter 2011, fall 2010 served as the baseline).[16] Incentivized students were offered one of the following rewards: financial low ($10 cash), financial high ($20 cash), or nonfinancial (trophy). The financial high and nonfinancial rewards were offered either in the gain frame or in the loss frame. In the loss conditions (financial high and nonfinancial) students received the reward at the start of the testing session, kept the reward at their computer during testing 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 indicated on the form what they planned to do with it. We also tested a delayed variant of the four most effective rewards: financial high, nonfinancial, financial loss, and nonfinancial loss. The delayed rewards were identical to the immediate rewards except that students were told they would receive the reward a month after testing.

As in Bloom and Chicago Heights, the test administrator held up the reward at the front of the room. Immediately after testing we privately informed students whether they had improved and distributed the rewards (except in delayed treatments, where this took place a month after testing). In the loss treatments, we collected the upfront incentives from all students at the end of testing and then privately returned rewards

[15] For reading, Scantron results have a 0.755–0.844 correlation with ISAT reading scores in grades four to eight. Math score correlations range from 0.749–0.823 (Davis 2010).

[16] In fall 2010, second graders were taking the test for the first time and therefore did not have a baseline score. They were offered a reward for scoring as high as the average second grader in the previous cohort.

to qualifying students. Redistribution occurred immediately after testing in immediate treatments and one month after testing in delayed treatments.

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). 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. In the results below, we pool the control groups: Control—Statement and Control—Statement Delayed in the 2010 wave; and Control—Statement and Control—No Statement in the 2011 wave. The groups do not differ in within wave test performance at the 10 percent significance level, and the results are robust to excluding individual control groups (Appendix B Table B3).

As noted above, students were not time constrained on the test. However, for about 15 percent of students the time reserved for the testing session ended before they completed the test. In these cases, students returned for a second session to complete the test and rewards were distributed immediately after the final testing session. The results are robust to excluding students who did not complete the test during the initial treatment session (Appendix B Table B3).

We randomized at the level of school-grade and blocked the randomization on school, grade, and average school-grade baseline math and reading scores. As in Bloom 2009, students who participated in the first CPS wave (2010) were re-randomized for the second wave (2011). In the second 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 reading intervention that took place between the two waves. The intervention, which incentivized students to read books, does not affect test performance and our results are robust to excluding students exposed to the intervention (Appendix B Table B3). As in Bloom, we also examine both spillovers to future testing and the results by session in order to address concerns about the effect of previous treatments on student responsiveness to our incentives. As discussed in more detail below, we find little impact of treatment on future test performance (Table 11). We do find differences in treatment effects across sessions (Appendix B Table B2) but, as also discussed below, this is not due to students participating in a prior session.

Table 5 reports summary statistics by treatment group for pretreatment characteristics in CPS pooling the 2010 and 2011 waves. The pretreatment characteristics include baseline score on the tested subject (either math or reading), grade, test subject, and the following demographics: gender, race/ethnicity, free or reduced price lunch status, and eligibility for an Individualized Education Plan (IEP). We standardize test scores within session, test subject, and school-grade to have mean zero and standard deviation one using the full population of CPS students, and cluster standard errors by school-grade. While the groups are generally balanced, the table indicates the presence of some significant differences between individual incentive treatments and control, as well as some imbalance in the overall distribution of students across treatments.

There are individually statistically significant differences (both positive and negative) in baseline test scores, the proportion of math tests, as well as demographic

Table 5—Baseline Characteristics by Treatment Group: Chicago Public Schools (CPS)

	Control	Immediate rewards					
		Financial low	Financial high	Non-financial	Financial loss	Non-financial loss	p-value
Panel A							
Observations	2,088	135	887	664	948	841	
Baseline test score	0.016	0.227	−0.028	0.000	−0.113	−0.010	0.154
	(0.890)	(0.878)	(0.909)	(0.921)	(0.930)	(0.937)	
Grade	5.255	5.556	5.202	5.111	4.811	4.925	0.917
	(1.862)	(1.336)	(1.789)	(1.949)	(2.000)	(1.879)	
Subject—Math	0.301	0.222	0.286	0.167	0.259	0.359	0.708
	(0.459)	(0.416)	(0.452)	(0.373)	(0.438)	(0.480)	
Female	0.534	0.622	0.532	0.505	0.563	0.486	0.036
	(0.499)	(0.485)	(0.499)	(0.500)	(0.496)	(0.500)	
Black	0.986	0.993	0.982	0.995	0.992	0.980	0.022
	(0.117)	(0.083)	(0.133)	(0.071)	(0.089)	(0.140)	
Free or reduced price lunch	0.984	1.000	0.981	0.983	0.976	0.983	0.843
	(0.125)	(0.000)	(0.137)	(0.129)	(0.153)	(0.129)	
Individualized Education Plan (IEP)	0.074	0.067	0.091	0.086	0.092	0.107	0.194
	(0.262)	(0.250)	(0.288)	(0.280)	(0.289)	(0.309)	
	Control		Delayed rewards				
			Financial high	Non-financial	Financial loss	Non-Financial loss	p-value
Panel B							
Observations	2,088		133	168	44	117	
Baseline test score	0.016		0.123	0.107	−0.009	0.202	0.154
	(0.890)		(0.907)	(0.991)	(0.957)	(1.003)	
Grade	5.255		5.150	4.583	5.023	4.547	0.917
	(1.862)		(1.323)	(1.639)	(1.677)	(1.873)	
Subject—Math	0.301		0.421	0.214	0.000	0.000	0.708
	(0.459)		(0.494)	(0.410)	(0.000)	(0.000)	
Female	0.534		0.515	0.464	0.349	0.530	0.036
	(0.499)		(0.500)	(0.499)	(0.477)	(0.499)	
Black	0.986		0.977	0.982	0.953	0.991	0.022
	(0.117)		(0.150)	(0.133)	(0.212)	(0.094)	
Free or reduced price lunch	0.984		0.977	0.964	0.977	0.991	0.843
	(0.125)		(0.150)	(0.186)	(0.150)	(0.094)	
Individualized Education Plan (IEP)	0.074		0.068	0.072	0.070	0.043	0.194
	(0.262)		(0.252)	(0.258)	(0.255)	(0.203)	

Notes: The table reports group means pooling the CPS 2010 and CPS 2011 waves. Standard deviations are reported in parentheses. Baseline score is standardized within grade, subject and testing period to have mean zero and standard deviation one using the full sample of CPS students. The p-value reports the probability from a joint F-test over both panels that the means are equal to one another, clustering by school-grade. Treatments that were completely homogeneous were not included in the F-test.

measures in some groups. In some treatments there is no within-treatment variation for certain variables. For example, in the financial low treatment, 100 percent of the sample receives free or reduced lunch; and, in both of the delayed loss treatments there are no math subject tests. In these cases, the implied standard deviation is zero, leading to a rejection of the null hypothesis of equal means, even when differences across treatments are small (e.g., free/reduced lunch eligibility proportions of 0.984

Table 6—Effects of Immediate Rewards on Test Performance

	Pooled		Bloom 2009		Bloom 2010	
	(1)	(2)	(3)	(4)	(5)	(6)
Financial low	0.009	−0.008	0.064	0.045		
	(0.042)	(0.041)	(0.079)	(0.071)		
Financial high	0.093	0.068	0.245	0.206	0.129	0.124
	(0.030)	(0.027)	(0.119)	(0.099)	(0.056)	(0.060)
Nonfinancial	0.044	0.040				
	(0.032)	(0.034)				
Financial loss	0.153	0.123			0.211	0.212
	(0.031)	(0.030)			(0.054)	(0.056)
Nonfinancial loss	0.115	0.097				
	(0.041)	(0.037)				
Additional covariates		Yes		Yes		Yes
Observations	6,843	6,843	584	584	333	333
Classes/school-grades	227	227	18	18	22	22
	Chicago Heights		CPS fall		CPS winter	
	(7)	(8)	(9)	(10)	(11)	(12)
Financial low	0.245	0.241	0.051	0.043		
	(0.070)	(0.070)	(0.081)	(0.075)		
Financial high	0.319	0.299	0.091	0.084	0.055	−0.003
	(0.155)	(0.177)	(0.063)	(0.050)	(0.036)	(0.035)
Nonfinancial	0.289	0.285	0.029	0.040	−0.001	−0.022
	(0.142)	(0.133)	(0.104)	(0.092)	(0.037)	(0.035)
Financial loss			0.233	0.209	0.066	0.010
			(0.065)	(0.062)	(0.035)	(0.033)
Nonfinancial loss			0.267	0.297	0.045	0.001
			(0.074)	(0.081)	(0.049)	(0.044)
Additional covariates		Yes		Yes		Yes
Observations	363	363	1,725	1,725	3,838	3,838
Classes/school-grades	17	17	89	89	165	165

Notes: The table reports OLS estimates for treatment effects on standardized test scores for each wave. Robust standard errors clustered by class in Bloom and by school-grade in Chicago Heights and CPS are reported in parentheses. The omitted category in each regression is the pooled control group in the relevant setting(s). All regressions include controls for the variables we block all randomizations on: session, school, grade, and baseline test score (score, score squared, score cubed). Even-numbered columns add controls for: past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status, where applicable.

compared to 1.0). We therefore exclude treatments that are completely homogeneous from the F-test of equal means across groups (reported in the final column). There is overall imbalance in the proportion female and the distribution of black students across treatments. As shown in Section II, the results are robust to including controls for pretreatment characteristics.

II. Results

Table 6 reports our basic results for all of our treatments in which the rewards were delivered immediately (as opposed to with a one month delay). We estimate treatment effects in both the pooled sample and for each wave in our individual settings: Bloom Township (Bloom), Chicago Heights (CH), and Chicago Public Schools (CPS). The dependent variable in all regressions is standardized test score with standard errors clustered by class (Bloom) or school-grade (CH and CPS). All regressions include controls for the variables we blocked the randomization on in all settings: session, school, grade, and baseline test score (score, score squared and score cubed).[17] Even-numbered columns add controls for past treatment, test subject, gender, race/ethnicity, free/reduced lunch eligibility, and (in CH and CPS) IEP status.[18] The omitted category in every regression is the pooled control (statement and no statement) group. There are no significant differences in performance between the control subgroups and pooling does not affect the results (Table B3). This suggests that the treatment effects are due to the incentives rather than the presence of the experimenters or the mere encouragement to improve.

Before proceeding to the overall results, we draw the reader's attention to the fact that four of our five test sessions yielded large and generally statistically significant impacts. In stark contrast, we find virtually no effects in the second wave of interventions conducted at CPS (the final two columns of Table 6). We have no compelling explanation for this discrepancy. It is not due to students receiving treatment for a second time, because the null result is also present for the large group of students treated for the first time in that wave. We have searched extensively for evidence of either a mistake in how we implemented that session or a mistake in our data recording and analysis, but have found neither.

A. *Result 1: Large and Immediate Monetary Incentives Lead to Test Score Improvements*, *Small Monetary Incentives Do Not*

The first result that emerges from Table 6 is the power of large and immediate financial incentives to increase test scores. The point estimates of the $20 incentives (framed either as a gain or a loss) are consistently positive and statistically significant at conventional levels, with improvements ranging from 0.068–0.153 standard deviations in the pooled sample.[19] 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. The magnitude of the impact is equivalent to about five months' worth of learning on the test.[20]

[17] We set all missing baseline test scores to zero. As noted above, all second graders in the CPS 2010 wave are missing baseline test scores.

[18] We include an indicator variable for missing covariates. Past treatment controls for the type of incentives received in previous testing sessions for Bloom spring 2009 and CPS 2011. In CPS 2011, past treatment also includes the type of treatment (if any) a student received in the separate reading intervention (discussed above) that took place between the two CPS waves.

[19] The estimates significant at the $p < 0.01$ level are robust to the Bonferroni correction for multiple hypothesis testing at a significance level of 5 percent. The correction procedure of List, Shaikh, and Xu (2016) is not applicable to our analysis because all our OLS estimates include covariates.

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

In contrast, however, we see little or no impact from the \$10 incentives, which are only effective in Chicago Heights. 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.[21] One interpretation is that, at least for some students, effort costs may be relatively high.[22] Together these results provide evidence that students understand the production function for this task but require sufficient motivation to exert effort.

B. *Result 2: Nonfinancial Incentives Also Impact Performance*

Turning to our first behavioral intervention, we compare the effects of non-pecuniary rewards to the effects of both low and high monetary rewards, which allows us to price out the effects of nonfinancial incentives. In the pooled results, the point estimates for non-pecuniary rewards (framed either as a gain or a loss) are somewhat smaller than those for the \$20 treatment and much larger than those from the \$10 treatment.

Typically, the material cost of nonfinancial incentives is low—in our case, one trophy cost approximately \$3. Hence, nonfinancial incentives are a potentially much more cost effective way of improving student performance than is paying cash. As we discussed above, 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 rewards.

C. *Result 3: Incentives Framed as Losses Appear to Outperform Those Framed as Gains*

Our second behavioral intervention built on the large literature demonstrating the power of framing for influencing choices, especially in the gain/loss space. The bottom two rows of Table 6 report the estimates for our "loss" treatments: one using a financial incentive, the other a prize. In the pooled estimates, the coefficients on losses are roughly twice the magnitude of the analogous "gain" treatments, but are not statistically different from those treatments. Thus, our results hint at the potential power of exploiting loss aversion in this context, but are not definitive.[23]

D. *Result 4: Rewards Provided with a Delay Have No Impact on Student Performance*

Perhaps the most striking and important finding of our study is that delayed rewards proved completely ineffective in raising test scores, as shown in Table 7. The structure of the table matches that of Table 6, except that the coefficients reported correspond to treatments in which the rewards were given to the students only after a one month delay and includes only the session and setting where they were tested (CPS 2010).

[21] In contrast, Barrow and Rouse (2013) find no evidence of sensitivity to reward size among post-secondary students offered semester-long incentives ranging from \$500 to \$1,000.

[22] It may also be the case that relatively low financial incentives crowd out intrinsic motivation yielding smaller net effects. We address this concern below.

[23] 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.

Table 7—Effects of Delayed Rewards

	CPS	
	(1)	(2)
Delayed financial high	−0.029	−0.050
	(0.104)	(0.103)
Delayed nonfinancial	0.033	0.042
	(0.046)	(0.052)
Delayed financial loss	0.181	0.222
	(0.123)	(0.150)
Delayed nonfinancial loss	−0.051	−0.005
	(0.100)	(0.100)
Additional covariates		Yes
Observations	2,052	2,052
Classes/school-grades	104	104

Notes: The table reports OLS estimates for treatment effects on standardized test scores for the CPS 2010 wave. Robust standard errors clustered by school-grade are reported in parentheses. The omitted category is the pooled control group. All regressions include controls for immediate incentive treatments (financial high, nonfinancial, financial loss, nonfinancial loss) and the variables we block the randomization on: school, grade, and baseline test score (score, score squared, score cubed). Column 2 adds controls for past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status.

All the regressions control for the analogous immediate incentive treatments. The coefficients on the delayed reward treatments are as likely to be negative as positive, and none are statistically significant. The only large, positive coefficients (delayed financial loss) are based on a small sample and thus carry large standard errors. The effects of the pooled delayed treatments are significantly different from the analogous pooled immediate treatments at the $p < 0.01$ level. The divergence between the immediate and delayed rewards reflect either hyperbolic discounting or enormously high exponential discount rates (i.e., over 800 percent annually).

While these findings are consistent with previous research highlighting the high discount rates of children, it poses a challenge for educators and policymakers. Typically, the results of statewide assessments are only available one to two months after the administration of the tests, making it difficult to provide immediate rewards for performance. 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.

In results 5–7 below, we investigate heterogeneous treatment effects. Tables 8, 9, and 10 report results for the immediate incentives split by age, test subject, and gender, respectively.[24] For space, we only present regressions that include the full set of covariate controls.[25] We estimate effects in each individual setting as well as in the pooled sample. The final column in panel A of each table reports p-values resulting

[24] We also examine treatment effects split by race/ethnicity (black and Hispanic) and baseline test score (below and above median) and find no evidence of differential treatment effects. Results are available upon request.

[25] Regressions that only include controls for the variables we block the randomization on yield similar results and are available upon request.

Table 8—Treatment Effects by Age

	Pooled		
	Elementary	Middle/secondary	p-value
Panel A			
Financial low	0.016	−0.012	0.999
	(0.072)	(0.053)	
Financial high	0.105	0.081	0.317
	(0.052)	(0.046)	
Nonfinancial	0.086	0.073	0.191
	(0.046)	(0.084)	
Financial loss	0.095	0.159	0.895
	(0.055)	(0.037)	
Nonfinancial loss	0.215	−0.073	0.021
	(0.048)	(0.047)	
Additional covariates	Yes	Yes	
Observations	3,335	3,508	
Classes/school-grades	106	121	

	Bloom	Chicago Heights		CPS	
	Secondary	Elementary	Middle	Elementary	Middle
Panel B					
Financial low	0.039	0.124	0.280	0.004	−0.085
	(0.064)	(0.060)	(0.069)	(0.096)	(0.082)
Financial high	0.178	0.444	−0.392	0.091	−0.011
	(0.068)	(0.149)	(0.089)	(0.055)	(0.057)
Nonfinancial		0.116	0.161	0.067	0.013
		(0.115)	(0.082)	(0.049)	(0.098)
Financial loss	0.259			0.097	0.097
	(0.069)			(0.052)	(0.038)
Nonfinancial loss				0.218	−0.115
				(0.049)	(0.045)
Additional covariates	Yes	Yes	Yes	Yes	Yes
Observations	917	179	184	3,156	2,407
Classes/school-grades	40	8	9	98	72

Notes: The table reports OLS estimates for treatment effects on standardized test scores for elementary (second–fifth grades), middle (sixth–eighth grades), and secondary (tenth grade) students in pooled waves in Bloom and CPS and a single wave in Chicago Heights. The reported p-values result from a test of equal coefficients for elementary and middle/secondary students in the pooled sample. Robust standard errors clustered by class in Bloom and by school-grade in Chicago Heights and CPS are reported in parentheses. The omitted category in each regression is the pooled control group in the relevant setting(s). All regressions include controls for session, school, grade, baseline test score (score, score squared, score cubed), past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status, where applicable.

from a test of equal coefficients across subgroups in the pooled sample. The sample sizes in Chicago Heights are quite small relative to the other sites (especially CPS), and thus are less stable and less precisely estimated.

E. *Result 5: Younger Students May Respond More to Nonfinancial Incentives*

Table 8 estimates treatment effects separately for secondary (tenth grade) students in Bloom, and for elementary (second–fifth grade) and middle (sixth–eighth

Table 9—Treatment Effects by Test Subject

	Pooled		
	Reading	Math	p-value
Panel A			
Financial low	−0.080	0.173	0.000
	(0.052)	(0.052)	
Financial high	0.052	0.246	0.020
	(0.032)	(0.080)	
Nonfinancial	0.050	0.081	0.992
	(0.041)	(0.077)	
Financial loss	0.102	0.299	0.082
	(0.032)	(0.100)	
Nonfinancial loss	0.111	0.032	0.350
	(0.045)	(0.064)	
Additional covariates	Yes	Yes	
Observations	4,908	1,935	
Classes/school-grades	179	93	

	Bloom	Chicago Heights	CPS	
	Reading	Math	Reading	Math
Panel B				
Financial low	0.039	0.241	−0.224	0.137
	(0.064)	(0.070)	(0.070)	(0.072)
Financial high	0.178	0.299	0.000	0.238
	(0.068)	(0.177)	(0.035)	(0.088)
Nonfinancial		0.285	0.022	0.030
		(0.133)	(0.042)	(0.083)
Financial loss	0.259		0.061	0.283
	(0.069)		(0.035)	(0.100)
Nonfinancial loss			0.077	0.025
			(0.044)	(0.065)
Additional covariates	Yes	Yes	Yes	Yes
Observations	917	363	3,991	1,572
Classes/school-grades	40	17	139	76

Notes: The table reports OLS estimates for treatment effects on standardized test scores for reading (Bloom and CPS) and math (Chicago Heights and CPS) for students in pooled waves in Bloom and CPS and a single wave in Chicago Heights. The reported p-values result from a test of equal coefficients for math and reading in the pooled sample. Robust standard errors clustered by class in Bloom and by school-grade in Chicago Heights and CPS are reported in parentheses. The omitted category in each regression is the pooled control group in the relevant setting(s). All regressions include controls for session, school, grade, baseline test score (score, score squared, score cubed), past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status, where applicable.

grade) school students in Chicago Heights and CPS.[26] The pooled sample estimates treatment effects separately for elementary (second–fifth grade) students and middle/secondary (sixth–eighth and tenth grade) students. In general, we see similar results across young and old students, with the exception of nonfinancial incentives framed as losses, where we find large positive effects on young students

[26] Due to small sample sizes, we are not able to include school and grade fixed effects for Chicago Heights students.

Table 10—Treatment Effects by Gender

	Pooled		
	Male	Female	p-value
Panel A			
Financial low	0.049	−0.078	0.054
	(0.057)	(0.049)	
Financial high	0.046	0.078	0.446
	(0.038)	(0.031)	
Nonfinancial	0.042	0.040	0.960
	(0.044)	(0.040)	
Financial loss	0.144	0.104	0.416
	(0.044)	(0.034)	
Nonfinancial loss	0.122	0.065	0.289
	(0.051)	(0.040)	
Additional covariates	Yes	Yes	
Observations	3,277	3,566	
Classes/school-grades	227	226	

	Bloom		Chicago Heights		CPS	
	Male	Female	Male	Female	Male	Female
Panel B						
Financial low	0.134	−0.067	0.228	0.231	−0.070	−0.152
	(0.093)	(0.081)	(0.071)	(0.086)	(0.088)	(0.064)
Financial high	0.178	0.163	0.165	0.399	0.003	0.038
	(0.083)	(0.079)	(0.125)	(0.248)	(0.042)	(0.034)
Nonfinancial			0.264	0.304	0.030	0.029
			(0.091)	(0.169)	(0.046)	(0.042)
Financial loss	0.351	0.132			0.105	0.094
	(0.093)	(0.095)			(0.046)	(0.036)
Nonfinancial loss					0.099	0.050
					(0.052)	(0.043)
Additional covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	474	443	189	174	2,614	2,949
Classes/school-grades	40	40	17	17	170	169

Notes: The table reports OLS estimates for treatment effects on standardized test scores for males and females in pooled waves in Bloom and CPS and a single wave in Chicago Heights. The reported p-values result from a test of equal coefficients for males and females in the pooled sample. Robust standard errors clustered by class in Bloom and by school-grade in Chicago Heights and CPS are reported in parentheses. The omitted category in each regression is the pooled control group in the relevant setting(s). All regressions include controls for session, school, grade, baseline test score (score, score squared, score cubed), past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status, where applicable.

and small negative impacts on older students.[27] It seems sensible that younger children would be more affected by noncash rewards: they are less familiar with cash, might receive higher utility from the type of prize we were offering, and are also more likely to overestimate the value of nonfinancial rewards (for example, one third grader announced her estimated value of the $3 trophy to be $20). Our

[27] The nonfinancial loss treatment was only carried out in CPS. The coefficients on that treatment vary between the pooled regression and the CPS-specific regressions because the coefficients on the other covariates in the regression differ between the pooled and CPS regressions, indirectly impacting the estimated treatment effects.

findings suggest that among children with a limited understanding of monetary returns, nonfinancial rewards can be particularly cost-effective at addressing underinvestment in education.

F. *Result 6: Math Scores Respond More Strongly Than Reading Scores*

Table 9 presents treatment effects on reading (Bloom and CPS) and on math (Chicago Heights and CPS) tests. The gains in math are larger for four of the five treatments offered. Pooling all math treatments and all reading treatments, the difference is highly statistically significant. The pattern of results are similar if we restrict ourselves to CPS, which is the only setting that included both math and reading tests. The most likely explanation for this result is that math scores are more sensitive to effort than reading. And, indeed, it is often the case that educational incentives have a greater impact on math than reading (e.g., Decker, Mayer, and Glazerman 2004; Rockoff 2004; Jacob 2005; Dobbie and Fryer Jr. 2011).

G. *Result 7: Suggestive Evidence That Boys Are More Responsive Than Girls*

Table 10 presents results separately for boys and girls. We generally see larger responses to our interventions for boys relative to girls (except in Chicago Heights where treatment effects are larger for girls). The biggest gaps emerge with low financial stakes and in the nonfinancial loss treatment. Our findings with respect to gender are consistent with a wealth of prior research that shows boys tend to be more sensitive to short-term incentives than girls, which may be due in part to gender differences in time preferences.[28]

H. *Result 8: The Introduction of Rewards Has No Clear Impact on Future Test Scores, Except Perhaps a Crowding Out Effect of Low Financial Incentives*

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.[29] However, on tasks where intrinsic motivation is already low or zero, external rewards are less likely to have such negative long-term effects.[30] It is also worth noting that several studies have tracked student performance after incentives are removed and find little evidence of crowd

[28] 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 2009; Angrist and Lavy 2009) and shorter term studies finding larger effects among boys, particularly in the context of competition (Gneezy, Niederle, and Rustichini 2003; Gneezy and Rustichini 2004). Attali, Neeman, and Schlosser (2011) find that performance differences on high and low stakes tests are larger for males than females. Bettinger and Slonim (2007) and Castillo et al. (2011) find that boys are more impatient than girls.

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

[30] For further discussion see reviews by, e.g., Eisenberger and Cameron (1996), Camerer and Hogarth (1999), Deci, Koestner, and Ryan (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.

TABLE 11—TREATMENT EFFECTS ON FUTURE TEST SCORES

	Same subject, subsequent session				Subsequent subject, same test session	
	Bloom		CPS		CPS	
	(1)	(2)	(3)	(4)	(5)	(6)
Financial low	−0.159	−0.192	−0.124	−0.154	−0.080	−0.125
	(0.114)	(0.111)	(0.110)	(0.111)	(0.084)	(0.065)
Financial high	0.033	0.019	−0.051	−0.057	−0.053	−0.011
	(0.141)	(0.145)	(0.035)	(0.034)	(0.051)	(0.054)
Nonfinancial			−0.020	−0.014	−0.063	−0.106
			(0.038)	(0.044)	(0.053)	(0.066)
Financial loss			0.034	0.021	−0.077	−0.028
			(0.038)	(0.037)	(0.059)	(0.051)
Nonfinancial loss			0.036	0.034	0.077	0.114
			(0.040)	(0.043)	(0.061)	(0.057)
Subsequent treatment	Yes	Yes	Yes	Yes		
Additional covariates		Yes		Yes		Yes
Observations	309	309	5,315	5,315	4,600	4,600
Classes/school-grades	15	15	170	170	165	165

Notes: The table reports OLS estimates for treatment effects on the treated test subject in the subsequent test session (Bloom 2009, CPS winter and spring 2011) and the subsequent subject in the same test session (CPS fall 2010 and winter 2011). 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 in the relevant setting. Columns 1–4 include controls for treatment (if any) on the subsequent test. All regressions include controls for the variables we block the randomization on: session, school, grade, and baseline test score (score, score squared, score cubed). Even-numbered columns add controls for past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status, where applicable.

out (see, e.g., Bettinger 2012; Barrera-Osorio et al. 2011; Kremer, Miguel, and Thornton 2009; Levitt, List, and Sadoff 2010).[31]

We similarly 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 nonfinancial incentives. Table 11 explores two different dimensions along which temporary incentives might distort future outcomes. We first report the impact of exposure to treatment today on test scores in the same subject, but when taking the exam in the next testing period, months later. The final two columns estimate the effect of the various treatments on test scores from a subsequent non-incentivized test in a different subject taken in the same testing period, i.e., just hours or days later. Any increase or decrease in scores on this test would come only from an altered level of effort exerted on the test.[32] The results are

[31] Additionally, Bettinger (2012) finds no evidence that a test performance incentive program erodes elementary school students' intrinsic motivation measured using student and teacher surveys. Similarly, Barrow and Rouse (2013) find that performance based scholarships have no negative impacts on internal motivation, interest, or enjoyment in learning.

[32] In columns 1–4, we regress the student's treatment on her standardized test score taken in the subsequent period, controlling for any subsequent treatments when necessary. In Bloom, we regress winter 2009 treatment on spring 2009 test score. In CPS, we regress fall 2010 treatment on winter 2011 score, and winter 2011 treatment on spring 2011 score in the same subject (winter 2011 serves as the baseline score for spring 2011). Columns 5 and 6 include students who received treatment on their first subject test taken in the testing period in CPS fall 2010 and winter 2011. Here, we regress math (reading) treatment on reading (math) score in the same period. Controls for past treatment include CPS fall 2010 treatment for CPS spring 2011 in column 4 and CPS winter 2011 in column 6;

similar across these two settings. Interestingly, for low financial incentives (which do not even improve student performance on the incentivized test), there appears to be a consistently large and negative spillover effect on the order of one-tenth of a standard deviation, as in Gneezy and Rustichini (2000a, b). These spillovers are statistically significant only in the final column, but are jointly, highly significant. This result points to a real risk: small financial incentives not only yield no immediate effort response, but seem to discourage effort on other tests as well. In contrast, bigger financial rewards and nonfinancial rewards yield a highly mixed set of estimates, roughly as likely to be positive as negative.

III. Conclusion

Most education policies will fail if students do not exert effort. Yet, surprisingly little is known about what motivates students to invest effort in school or the causal impact of this effort on learning and achievement.[33] This is in part because in most educational settings, it is very difficult to disentangle student effort from student ability. For example, if a student performs badly on a test, is it because she does not understand the material or because she was not motivated to answer the questions correctly?

At the same time, the standard model—in which individuals choose their educational attainment based on the returns to schooling—does not fully capture the kinds of daily investments students must make in order to accumulate human capital. Many of the tasks that students perform (such as completing homework assignments, paying attention in class, etc.) are low stakes and yield benefits only far in the future. And it is the rare third grader who turns in her homework because of the marginal impact this will have on her (discounted) returns to schooling.

Instead, these policies seem to implicitly rely on other factors to drive student effort, including: intrinsic motivation, habit, norms, and extrinsic rewards provided, for example, by parents through explicit incentives, positive feedback, punishment, and praise. In contexts where these factors are not in place,[34] there is growing interest in the role of short-term incentives to increase student effort.

This study examines, in one particular context—effort exerted on low stakes tests—whether approaches suggested by behavioral economics can increase the effectiveness of such incentives. Our most striking finding relates to the sensitivity of students to the timing of rewards. We obtain large test score impacts when payments are made immediately, but no impact when rewards are delivered with a one-month delay. Given the long delay in most returns to education, these results could be consistent with a broad pattern of underinvestment in human capital by students. Further, we find an impact of nonfinancial rewards, especially for younger

and the type of treatment (if any) a student received in the separate reading intervention that took place between the two CPS waves for CPS spring 2011 in column 4 and CPS winter 2011 in columns 4 and 6.

[33] Barrow and Rouse (2013) measure effort responses to performance-based incentives for post-secondary students, and also discuss the dearth of evidence on the impact of student effort on achievement.

[34] We believe this is likely to be the case among many of the disadvantaged students in our study. Low-income parents are less likely than affluent parents to offer their children incentives for effort and achievement (Gottfried, Fleming, and Gottfried 1998). And these students are primarily located in low-educated neighborhoods and low-performing schools where their experience and the social norms may not conform to a model of high effort and high achievement (e.g., Wilson 1987; Austen-Smith and Fryer Jr. 2005).

students. Framing the rewards as losses may also increase their effectiveness. More broadly we demonstrate that on our low stakes task many students are investing little effort, and that effort alone can have a large impact on performance.[35]

We argue that motivating student effort is a critical and not well understood first step to crafting policies aimed at increasing achievement. With this goal in mind, an important limitation on the generalizability of our study is that we do not know how students would respond if these incentives were offered on a regular basis in order to motivate sustained effort in schooling, or whether repeated incentives would be cost effective. A next step in this research is to understand whether these kinds of incentives can be used to promote habit formation and learning.

While there is concern that incentives of the kind we examine will crowd out intrinsic motivation, we find little evidence for this to be true. However, we note that intrinsic motivation on our task is likely low at baseline. Our results suggest that the kinds of incentives we have designed will be most effective in contexts where students lack motivation on low stakes tasks.[36] In such cases, there is the notion that extrinsic rewards can actually be used to foster intrinsic motivation and habit formation (Cameron et al. 2005; Pierce et al. 2003; Bettinger 2012).

This can occur through several channels. If immediate rewards increase students' estimated utility returns to education, then properly structured extrinsic rewards could potentially build (rather than crowd out) intrinsic motivation. Similarly, students may learn that they enjoy exerting effort and hence learning more. If this occurs at the class or school level, it can potentially shift social norms around educational investments—e.g., behaving in class, wanting to get good grades, etc.[37] Short-term rewards can also address problems related to planning failures and limited understanding of the production function. Students may not know the steps to take in order to improve their achievement on a test that is six months away. However, they may be able to effectively respond to performance-based incentives on interim tasks such as learning the daily lesson, completing an assignment, or focusing on a practice test.

Finally, the kinds of incentives we study can build habits that carry forward even after the rewards are removed. Developing these habits may be an important skill in itself. Increasingly, psychologists and economists are demonstrating the importance of non-cognitive abilities such as self-control, persistence, conscientiousness, and grit in educational achievement and work success (e.g., Mischel, Shoda, and Rodriguez 1989; Duckworth and Seligman 2005; Duckworth et al. 2007; Heckman, Stixrud, and Urzua 2006). These traits are all characterized by a willingness to invest effort into activities that are low stakes in the near term but that contribute to a longer term goal. For students who lack motivation, occasional immediate rewards

[35] Metcalfe, Burgess, and Proud (2011) demonstrate a similar finding in a high stakes context.

[36] It remains an empirical question how our rewards would affect performance on tasks where baseline incentives or motivation is high. We might see no effect since there is little room to move effort, or possibly negative effects for example due to crowding out of intrinsic motivation or choking under the pressure of overly high stakes (e.g., Beilock 2010). In our study, there is no evidence of choking—i.e., that higher incentives reduce performance. As discussed in the results section, students were generally more responsive to larger incentives.

[37] See Bursztyn and Jensen (2015) for recent evidence on the influence of classroom and peer norms on individual investment in education.

applied to a wide number of low stakes tasks could induce them to exert effort in ways that help develop critical non-cognitive abilities (Eisenberger 1992).

This area of research requires further exploration before it can answer all of the policy questions of interest (Lavecchia, Liu, and Oreopoulos 2014). Our study is one step in this direction. Jalava, Joensen, and Pellas (2014), which explores the impact of a range of different nonfinancial incentives in a similar low stakes testing environment, represents further progress in this direction. Future interventions can build on these findings to help educators identify when students may lack motivation and how best to increase student engagement and effort. More generally, continuing to apply important elements of behavioral economics to issues within education can directly aid practitioners in need of fresh approaches to the urban school problem. Such behavioral insights can strengthen the impact and the cost effectiveness of interventions in education. They 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 are in a unique position to examine theories heretofore untestable.

Appendix A: Administrator Scripts

A. *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.

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.

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. Please try to improve your score from 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.

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.

B. *Chicago Heights*

Common to All Treatments.—

To the teacher:

Please read the following statement to your students immediately before they begin the ThinkLink 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 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.

Nonfinancial (**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 TROPHY AND PHOTO*.] You will receive the trophy and be photographed immediately after the test.

C. *Chicago Public Schools* (*CPS*)

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):

CPS 2010.—

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.

Nonfinancial (**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.

Nonfinancial (**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.

Nonfinancial 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.

Nonfinancial 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.

CPS 2011.—

Control—Statement: You are about to take the Scantron test. You also took Scantron in the fall. Please try to improve your score from the fall. You will learn your score immediately after the test.

Financial High (**$20**)**:** You are about to take the Scantron test. You also took Scantron in the fall. Please try to improve your score from the fall. If you improve your score from the fall, you will receive \$20. You will learn your score and be paid in cash immediately after the test.

Financial Loss (\$20): You are about to take the Scantron test. You also took Scantron in the fall. Please try to improve your score from the fall.

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 fall, you will get to keep your \$20. If you do not improve your score from the fall, 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): You are about to take the Scantron test. You also took Scantron in the fall. Please try to improve your score from the fall. If you improve your score from the fall, you will receive this trophy [*SHOW SAMPLE TROPHY*]. You will learn your score and receive the trophy immediately after the test.

Nonfinancial Loss (Trophy): You are about to take the Scantron test. You also took Scantron in the fall. Please try to improve your score from the fall.

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

Appendix B: Tables

Table B1—Treatment Effects by Session: Bloom

			Wave 1 (2009)				Wave 2 (2010)	
	Pooled		Winter		Spring		Spring	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Financial low	0.052	0.039	0.096	0.075	−0.049	−0.080		
	(0.071)	(0.064)	(0.063)	(0.053)	(0.126)	(0.128)		
Financial high	0.212	0.178	0.233	0.213	0.232	0.179	0.129	0.124
	(0.080)	(0.068)	(0.112)	(0.109)	(0.170)	(0.150)	(0.056)	(0.060)
Financial loss	0.269	0.259					0.211	0.212
	(0.074)	(0.069)					(0.054)	(0.056)
Additional covariates		Yes		Yes		Yes		Yes
Observations	917	917	321	321	263	263	333	333
Classes	40	40	15	15	13	13	22	22

Notes: The table reports OLS estimates for treatment effects on standardized test scores in Bloom. Robust standard errors clustered by class are reported in parentheses. The omitted category in each regression is the pooled control group for the relevant session(s). All regressions include controls for baseline test score (score, score squared, score cubed). Column 1 includes controls for session. Even-numbered columns add controls for past treatment, gender, race/ethnicity, and free/reduced lunch status.

Table B2—Treatment Effects by Session: CPS

	Pooled		Fall		Winter	
	(1)	(2)	(3)	(4)	(5)	(6)
Financial low	−0.051	−0.103	0.051	0.043		
	(0.056)	(0.059)	(0.081)	(0.075)		
Financial high	0.056	0.026	0.091	0.084	0.055	−0.003
	(0.032)	(0.030)	(0.063)	(0.050)	(0.036)	(0.035)
Nonfinancial	0.028	0.028	0.029	0.040	−0.001	−0.022
	(0.033)	(0.035)	(0.104)	(0.092)	(0.037)	(0.035)
Financial loss	0.132	0.098	0.233	0.209	0.066	0.010
	(0.033)	(0.032)	(0.065)	(0.062)	(0.035)	(0.033)
Nonfinancial loss	0.100	0.079	0.267	0.297	0.045	0.001
	(0.041)	(0.039)	(0.074)	(0.081)	(0.049)	(0.044)
Additional covariates		Yes		Yes		Yes
Observations	5,563	5,563	1,725	1,725	3,838	3,838
School-grades	170	170	89	89	165	165

Notes: The table reports OLS estimates for treatment effects on standardized test scores in CPS. 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). All regressions include controls for the variables we block the randomization on: school, grade, and baseline test score (score, score squared, score cubed). Column 1 includes controls for session. Even-numbered columns add controls for past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status.

Table B3—Robustness Checks

	Chicago Heights	
		Separated from pooled control
	All students (1)	Control statement comparison (2)
Panel A		
Financial low	0.241	0.251
	(0.070)	(0.089)
Financial high	0.299	0.302
	(0.177)	(0.176)
Nonfinancial	0.285	0.285
	(0.133)	(0.132)
Control—statement comp.		0.051
		(0.136)
Additional covariates	Yes	Yes
Observations	363	363
School-grades	17	17

(*continued*)

Table B3—Robustness Checks (*continued*)

	Chicago Public Schools (CPS)				
				Separated from pooled control	
	All students (3)	Finished testing on time (4)	No exposure to reading intervention (5)	Control statement delayed (6)	Control no statement (7)
Panel B					
Financial low	−0.103	−0.102	−0.086	−0.142	−0.090
	(0.059)	(0.060)	(0.059)	(0.069)	(0.059)
Financial high	0.026	0.019	0.074	0.009	0.050
	(0.030)	(0.031)	(0.039)	(0.032)	(0.035)
Nonfinancial	0.028	0.021	0.032	0.010	0.050
	(0.035)	(0.038)	(0.045)	(0.037)	(0.038)
Financial loss	0.098	0.082	0.149	0.084	0.122
	(0.032)	(0.033)	(0.039)	(0.033)	(0.035)
Nonfinancial loss	0.079	0.102	0.105	0.064	0.102
	(0.039)	(0.047)	(0.052)	(0.041)	(0.043)
Control—statement delayed				−0.094	
				(0.055)	
Control—no statement					0.048
					(0.036)
Additional covariates	Yes	Yes	Yes	Yes	Yes
Observations	5,563	4,218	3,704	5,563	5,563
School-grades	170	157	122	170	170

Notes: The table reports OLS estimates for treatment effects on standardized test scores for pooled waves in CPS and a single wave in Chicago Heights. Robust standard errors clustered by school-grade are reported in parentheses. The omitted category in columns 1, 3, 4, and 5 is the pooled control group in the relevant setting. Columns 2, 6, and 7 each exclude one control condition from the baseline category by separately including it in the regression. Column 4 excludes students who did not complete testing in the treatment session. Column 5 excludes students in CPS winter 2011 who participated in the separate reading intervention in fall 2010. All regressions include controls for session, school, grade, baseline test score (score, score squared, score cubed), past treatment, test subject, gender, race/ethnicity, free/reduced lunch status, and IEP status, where applicable.

REFERENCES

Angrist, Joshua, 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): 847–62.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. “Incentives and Services for College Achievement: Evidence from a Randomized Trial.” *American Economic Journal: Applied Economics* 1 (1): 136–63.

Angrist, Joshua, and Victor Lavy. 2009. “The Effects 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): 544–55.

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack. 2014. “No margin, no mission? A field experiment on incentives for public service delivery.” *Journal of Public Economics* 120: 1–17.

Attali, Yigal, Zvika Neeman, and Analia Schlosser. 2011. “Rise to the Challenge or Not Give a Damn: Differential Performance in High vs. Low Stakes Tests.” Institute for the Study of Labor (IZA) Discussion Paper 5693.

Austen-Smith, David, and Roland G. Fryer, Jr. 2005. “An Economic Analysis of ‘Acting White.’” *Quarterly Journal of Economics* 120 (2): 551–83.

Ball, Sheryl, Catherine Eckel, Philip J. Grossman, and William Zame. 2001. “Status in Markets.” *Quarterly Journal of Economics* 116 (1): 161–88.

Barankay, Iwan. 2011. "Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment." http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.7793&rep=rep1&type=pdf.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2011. "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics* 3 (2): 167–95.

Barrow, Lisa, and Cecilia E. Rouse. 2013. "Financial Incentives and Educational Investment: The Impact of Performance-Based Scholarships on Student Time Use." National Bureau of Economic Research (NBER) Working Paper 19351.

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): 441–62.

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): 237–75.

Beilock, Sian. 2010. *Choke: What the Secrets of the Brain Reveal about Getting It Right When You Have to*. New York: Free Press.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78.

Bergman, Peter. 2012. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." http://www.columbia.edu/~psb2101/BergmanSubmission.pdf.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment." *Quarterly Journal of Economics* 125 (1): 263–306.

Bettinger, Eric, and Robert Slonim. 2007. "Patience among children." *Journal of Public Economics* 91 (1): 343–63.

Bettinger, Eric P. 2012. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." *Review of Economics and Statistics* 94 (3): 686–98.

Blanes i Vidal, Jordi, and Mareike Nossol. 2011. "Tournaments Without Prizes: Evidence from Personnel Records." *Management Science* 57 (10): 1721–36.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non. Forthcoming. "Employee Recognition and Performance: A Field Experiment." *Management Science*.

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* 113 (11): 2309–44.

Bursztyn, Leonardo, and Robert Jensen. 2015. "How Does Peer Pressure Affect Educational Investments?" *Quarterly Journal of Economics* 130 (3): 1329–67.

Cadena, Brian C., and Benjamin J. Keys. 2015. "Human Capital and the Lifetime Costs of Impatience." *American Economic Journal: Economic Policy* 7 (3): 126–53.

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*. Westport, CT: Bergin and Garvey.

Cameron, Judy, W. David Pierce, Katherine M. Banko, and Amber Gear. 2005. "Achievement-Based Rewards and Intrinsic Motivation: A Test of Cognitive Mediators." *Journal of Educational Psychology* 97 (4): 641–55.

Castillo, Marco, Paul J. 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): 1377–85.

Davis, Jim. 2010. "Review of Scantron Performance Series." http://www.gocatgo.com/texts/esr505.davis.instrument.review.pdf.

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): 627–68.

Decker, Paul T., Daniel P. Mayer, and Steven Glazerman. 2004. "The Effects of Teach for America on Students: Findings from a National Evaluation." Mathematica Policy Research (MPR) Paper 8792-750.

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

Discovery Education. 2008. "Discovery Education Assessment Research." http://www.discoveryeducation.com/pdf/assessment/Discovery_Education_Assessment_Research.pdf.

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

Duckworth, Angela L., and Martin E. P. Seligman. 2005. "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents." *Psychological Science* 16 (12): 939–44.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. "Grit: Perseverance and passion for long-term goals." *Journal of Personality and Social Psychology* 92 (6): 1087–1101.

Dynarski, Susan M., and Judith E. Scott-Clayton. 2008. "Complexity and Targeting in Federal Student Aid: A Quantitative Analysis." In *Tax Policy and the Economy*, Vol. 22, edited by James M. Poterba, 109–50. Chicago: University of Chicago Press.

Eisenberger, Robert. 1992. "Learned Industriousness." *Psychological Review* 99 (2): 248–67.

Eisenberger, Robert, and Judy Cameron. 1996. "Detrimental effects of reward: Reality or myth?" *American Psychologist* 51 (11): 1153–66.

Frey, Bruno S. 2007. "Awards as compensation." *European Management Review* 4 (1): 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* 97 (4): 746–55.

Fryer, Roland G., Jr. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *Quarterly Journal of Economics* 126 (4): 1755–98.

Fryer, Roland G., Jr. Steven D., Levitt, John List, and Sally Sadoff. 2012. "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment." National Bureau of Economic Research (NBER) Working Paper 18237.

Ganzach, Yoav, and Nili Karsahi. 1995. "Message framing and buying behavior: A field experiment." *Journal of Business Research* 32 (1): 11–17.

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

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

Gneezy, Uri, and Aldo Rustichini. 2004. "Gender and Competition at a Young Age." *American Economic Review* 94 (2): 377–81.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118 (3): 1049–74.

Gottfried, Adele Eskeles, James S. Fleming, and Allen W. Gottfried. 1998. "Role of Cognitively Stimulating Home Environment in Children's Academic Intrinsic Motivation: A Longitudinal Study." *Child Development* 69 (5): 1448–60.

Gruber, Jonathan. 2001. *Risky Behavior among Youths: An Economic Analysis*. Chicago: University of Chicago Press.

Hanemann, W. Michael. 2003. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 93 (1): 635–47.

Hastings, Justine S., and Olivia S. Mitchell. 2011. "How Financial Literacy and Impatience Shape Retirement Wealth and Investment Behaviors." National Bureau of Economic Research (NBER) Working Paper 16740.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123 (4): 1373–1414.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3): 411–82.

Hossain, Tanjim, and John A. List. 2012. "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." *Management Science* 58 (12): 2151–67.

Hoxby, Caroline, and Sarah Turner. 2013. "Expanding College Opportunities for High-Achieving, Low Income Students." Stanford Institute for Economic Policy Research (SIEPR) Discussion Paper 12-014.

Huberman, Bernardo A., Christoph H. Loch, and Ayse Önçüler. 2004. "Status As a Valued Resource." *Social Psychology Quarterly* 67 (1): 103–14.

Jacob, Brian A. 2005. "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics* 89 (5–6): 761–96.

Jalava, Nina, Juanna Schrøter Joensen, and Elin Pellas. 2014. "Grades and rank: Impacts of non-financial incentives on test performance." *Journal of Economic Behavior and Organization* 115: 161–96.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125 (2): 515–48.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–92.

Kirby, Kris N., Ricardo Godoy, Victoria Reyes-García, Elizabeth Byron, Lilian Apaza, William R. Leonard, Eddy Pérez, et al. 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): 213–22.

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

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

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *Review of Economics and Statistics* 91 (3): 437–56.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2): 443–77.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos. 2014. "Behavioral Economics of Education: Progress and Possibilities." National Bureau of Economic Research (NBER) Working Paper 20609.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. 2010. "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment." *Journal of the European Economic Association* 8 (6): 1243–65.

Levitt, Steven D., John A. List, and Sally Sadoff. 2016. "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment." National Bureau of Economic Research (NBER) Working Paper 22107.

Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff. 2016. "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance: Dataset." *American Economic Journal: Economic Policy*. http://dx.doi.org/10.1257/pol.20130358.

List, John A., and Anya Savikhin Samek. 2015. "The behavioralist as nutritionist: Leveraging behavioral economics to improve child food choice and consumption." *Journal of Health Economics* 39: 135–46.

List, John A., Azeem M. Shaikh, and Yang Xu. 2016. "Multiple Hypothesis Testing in Experimental Economics." National Bureau of Economic Research (NBER) Working Paper 21875.

Loewenstein, George, and Samuel Issacharoff. 1994. "Source dependence in the valuation of objects." *Journal of Behavioral Decision Making* 7 (3): 157–68.

Metcalfe, Robert, Simon Burgess, and Steven Proud. 2011. "Student effort and educational attainment: Using the England football team to identify the education production function." Centre for Market and Public Organisation (CMPO) Working Paper 11/276.

Mischel, Walter, Yuichi Shoda, and Monica I. Rodriguez. 1989. "Delay of gratification in children." *Science* 244 (4907): 933–38.

Nguyen, Trang. 2008. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." https://www.povertyactionlab.org/sites/default/files/documents/Nguyen%202008.pdf.

O'Neil, Harold F., Jamal Abedi, Judy Miyoshi, and Ann Mastergeorge. 2005. "Monetary Incentives for Low-Stakes Tests." *Educational Assessment* 10 (3): 185–208.

O'Neil, Harold F., Jr., Brenda Sugrue, and Eva L. Baker. 1995. "Effects of Motivational Interventions on the National Assessment of Educational Progress Mathematics Performance." *Educational Assessment* 3 (2): 135–57.

Oreopoulos, Philip. 2007. "Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling." *Journal of Public Economics* 91 (11): 2213–29.

Patel, Reshma, Lashawn Richburg-Hayes, Elijah de la Campa, and Timothy Rudd. 2013. *Performance-Based Scholarships: What Have We Learned? Interim Findings from the PBS Demonstration*. MDRC. New York, August.

Pierce, W. David, Judy Cameron, Katherine M. Banko, and Sylvia So. 2003. "Positive effects of rewards and performance standards on intrinsic motivation." *Psychological Record* 53 (4): 561–79.

Renaissance Learning. 2015. "STAR Reading Technical Manual." http://doc.renlearn.com/KMNet/R004384910GJF6AC.pdf.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94 (2): 247–52.

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.

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 Henry. 1955. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies* 23 (3): 165–80.

Thaler, Richard. 1980. "Toward a positive theory of consumer choice." *Journal of Economic Behavior and Organization* 1 (1): 39–60.

Tran, Anh, and Richard Zeckhauser. 2012. "Rank as an inherent incentive: Evidence from a field experiment." *Journal of Public Economics* 96 (9): 645–50.

Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein. 2008. "Financial Incentive-Based Approaches for Weight Loss." *Journal of the American Medical Association* 300 (22): 2631–37.

Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: University of Chicago Press.

This article has been cited by:

1. Zahra Sharafi. 2023. Poverty and perseverance: The detrimental effect of poverty on effort provision. *Journal of Development Economics* **162**, 103040. [Crossref]
2. Joshua J. Lewer, Colin Corbett, Tanya M. Marcum, Jannett Highfill. 2023. Offering makeup exams: do teachers care 'what happened'?. *Applied Economics Letters* **30**:5, 645-648. [Crossref]
3. Yukari Yamada, Tomoe Uchida, Shusaku Sasaki, Masataka Taguri, Takayuki Shiose, Tatsuyoshi Ikenoue, Shingo Fukuma. 2023. Nudge-Based Interventions on Health Promotion Activity Among Very Old People: A Pragmatic, 2-Arm, Participant-Blinded Randomized Controlled Trial. *Journal of the American Medical Directors Association* **24**:3, 390-394.e5. [Crossref]
4. Elif Incekara-Hafalir, Grace H. Y. Lee, Audrey K. L. Siah, Erte Xiao. 2023. Incentives to Persevere. *Management Science* **134**. . [Crossref]
5. Neal H. Olitsky, Sarah B. Cosgrove. 2023. Cutting our losses: The effects of a loss-aversion strategy on student learning gains. *The Journal of Economic Education* **54**:1, 1-16. [Crossref]
6. Nagaraju Dasari, Mohiddin Shaw Shaik, Mahendra Parihar, Indu Priyanka Dasari. A Sensible Solution to an Unintended Consequence of Relative Grading 15-23. [Crossref]
7. Matthew G. Springer. 2023. You Get What You Pay For: Why We Need to Invest in Strategic Compensation Reform. *Peabody Journal of Education* **98**:1, 66-82. [Crossref]
8. Roland G. Fryer Jr., Steven D. Levitt, John List, Sally Sadoff. 2022. Enhancing the Efficacy of Teacher Incentives through Framing: A Field Experiment. *American Economic Journal: Economic Policy* **14**:4, 269-299. [Abstract] [View PDF article] [PDF with links]
9. Paul J. Ferraro, J. Dustin Tracy. 2022. A reassessment of the potential for loss-framed incentive contracts to increase productivity: a meta-analysis and a real-effort experiment. *Experimental Economics* **25**:5, 1441-1466. [Crossref]
10. Lukas Kiessling, Jonas Radbruch, Sebastian Schaube. 2022. Self-Selection of Peers and Performance. *Management Science* **68**:11, 8184-8201. [Crossref]
11. Chiradip Chatterjee, Nafisa Halim, Pallab Mozumder. 2022. Energy conservation and health risk reduction: an experimental investigation of punishing vs. rewarding incentives. *Environmental Economics and Policy Studies* **24**:4, 551-570. [Crossref]
12. Eszter Czibor, Danny Hsu, David Jimenez-Gomez, Susanne Neckermann, Burcu Subasi. 2022. Loss-Framed Incentives and Employee (Mis-)Behavior. *Management Science* **68**:10, 7518-7537. [Crossref]
13. Puneet Arora, Nicholas Wright. 2022. Grade reporting and student performance. *Education Economics* **30**:4, 356-363. [Crossref]
14. Xiaofei Pan, Sukki Yoon. 2022. Gym Membership Programs: Image Motivation and Conditional Discount Framing. *Journal of Current Issues & Research in Advertising* **43**:3, 301-318. [Crossref]
15. Yunlong JIAN, Yuan LIU. 2022. Structure and effects of motivation: From the perspective of the motivation continuum. *Advances in Psychological Science* **30**:7, 1589-1603. [Crossref]
16. Maninder Singh, P.S. James, Happy Paul, Kartikeya Bolar. 2022. Impact of cognitive-behavioral motivation on student engagement. *Heliyon* **8**:7, e09843. [Crossref]
17. Jessica F. Schwab, Leah H. Somerville. 2022. Raising the Stakes for Online Learning: Monetary Incentives Increase Performance in a Computer-Based Learning Task Under Certain Conditions. *Frontiers in Psychology* **13**. . [Crossref]

18. Michael Yao-Ping Peng, Zizai Zhang. 2022. Future Time Orientation and Learning Engagement Through the Lens of Self-Determination Theory for Freshman: Evidence From Cross-Lagged Analysis. *Frontiers in Psychology* **12**. . [Crossref]
19. Claire Duquennois. 2022. Fictional Money, Real Costs: Impacts of Financial Salience on Disadvantaged Students. *American Economic Review* **112**:3, 798-826. [Abstract] [View PDF article] [PDF with links]
20. Heather J. Nuske, Jacqueline E. Buck, Brinda Ramesh, Emily M. Becker-Haimes, Kelly Zentgraf, David S. Mandell. 2022. Making Progress Monitoring Easier and More Motivating: Developing a Client Data Collection App Incorporating User-Centered Design and Behavioral Economics Insights. *Social Sciences* **11**:3, 106. [Crossref]
21. Gerhard Riener, Valentin Wagner. 2022. Non-monetary rewards in education. *Educational Psychology* **42**:2, 222-239. [Crossref]
22. Emily Lyons, Almaz Mesghina, Lindsey E. Richland. 2022. Complicated Gender Gaps in Mathematics Achievement: Elevated Stakes during Performance as One Explanation. *Mind, Brain, and Education* **16**:1, 36-47. [Crossref]
23. Tom McNamara, Debrah Meloso, Marco Michelotti, Petya Puncheva-Michelotti. 2022. 'You are free to choose . . . are you?' Organisational punishment as a productivity incentive in the social science literature. *Human Relations* **75**:2, 322-348. [Crossref]
24. Fritz Schiltz, Kristof De Witte. 2022. Sugar rush or sugar crash? Experimental evidence on the impact of sugary drinks in the classroom. *Health Economics* **31**:1, 215-232. [Crossref]
25. Sofiya Kobylyanskaya. Speech and Eye Tracking Features for L2 Acquisition: A Multimodal Experiment 47-52. [Crossref]
26. Linda Bol, Monica Christina Esqueda, Diane Ryan, Sue C. Kimmel. 2022. A Comparison of Academic Outcomes in Courses Taught With Open Educational Resources and Publisher Content. *Educational Researcher* **51**:1, 17-26. [Crossref]
27. Ali Bhayani. 2021. Let students pay for their higher education: Debate concerning free and subsidized education based on sunk cost theory. *Journal of Philanthropy and Marketing* **26**:4. . [Crossref]
28. Joshua L. Howard, Julien Bureau, Frédéric Guay, Jane X. Y. Chong, Richard M. Ryan. 2021. Student Motivation and Associated Outcomes: A Meta-Analysis From Self-Determination Theory. *Perspectives on Psychological Science* **16**:6, 1300-1323. [Crossref]
29. Thomas P. Andrews. 2021. "Provide a complete, concise economic analysis of the following article...": Using outside readings to train students to answer a single question. *The Journal of Economic Education* **52**:4, 316-325. [Crossref]
30. Chiradip Chatterjee, Nafisa Halim, Pallab Mozumder. 2021. Emission Tax, Health Insurance, and Information: A Mechanism Design for Reducing Energy Consumption and Emission Risk. *Economics of Disasters and Climate Change* **5**:3, 465-480. [Crossref]
31. Joshua J. Lewer, Colin Corbett, Tanya M. Marcum, Jannett Highfill. 2021. Modeling Student Effort: Flat Tires and Dead Batteries. *The American Economist* **66**:2, 301-314. [Crossref]
32. Yan Chen, Peter Cramton, John A. List, Axel Ockenfels. 2021. Market Design, Human Behavior, and Management. *Management Science* **67**:9, 5317-5348. [Crossref]
33. David Willinger, Iliana I. Karipidis, Plamina Dimanova, Susanne Walitza, Silvia Brem. 2021. Neurodevelopment of the incentive network facilitates motivated behaviour from adolescence to adulthood. *NeuroImage* **237**, 118186. [Crossref]

34. Christopher Roby. 2021. Can loss framing improve coordination in the minimum effort game?. *Journal of Economic Interaction and Coordination* **16**:3, 557-588. [Crossref]

35. Paul E. Carrillo, Edgar Castro, Carlos Scartascini. 2021. Public good provision and property tax compliance: Evidence from a natural experiment. *Journal of Public Economics* **198**, 104422. [Crossref]

36. Yahya İLTÜZER, Yasemin DEMİRASLAN ÇEVİK. 2021. ÇEVRİMİÇİ ÖĞRENME ORTAMLARINDA KULLANILAN DÜRTME STRATEJİLERİNİN ÜNİVERSİTE ÖĞRENCİLERİNİN PERFORMANSLARINA ETKİSİ VE PERFORMANS İLE MOTİVASYONLARINA YÖNELİK GÖRÜŞLERİ. *Eğitim Teknolojisi Kuram ve Uygulama* . [Crossref]

37. Setareh Maghsudi, Andrew Lan, Jie Xu, Mihaela van der Schaar. 2021. Personalized Education in the Artificial Intelligence Era: What to Expect Next. *IEEE Signal Processing Magazine* **38**:3, 37-50. [Crossref]

38. Nagaraju Dasari, Mahendra Parihar, Mohiddin Shaw Shaik. 2021. Analysis of Exertion with Relative Grading Mechanism in Academics Using Computational Technique. *IOP Conference Series: Materials Science and Engineering* **1099**:1, 012018. [Crossref]

39. Ildo Lautharte, Victor Hugo de Oliveira, Andre Loureiro. Incentives for Mayors to Improve Learning: Evidence from State Reforms in Ceará, Brazil **17**, . [Crossref]

40. Maya Escueta, Andre Joshua Nickow, Philip Oreopoulos, Vincent Quan. 2020. Upgrading Education with Technology: Insights from Experimental Research. *Journal of Economic Literature* **58**:4, 897-996. [Abstract] [View PDF article] [PDF with links]

41. Damon Clark, David Gill, Victoria Prowse, Mark Rush. 2020. Using Goals to Motivate College Students: Theory and Evidence From Field Experiments. *The Review of Economics and Statistics* **102**:4, 648-663. [Crossref]

42. Hanna M. Sittenthaler, Alwine Mohnen. 2020. Cash, non-cash, or mix? Gender matters! The impact of monetary, non-monetary, and mixed incentives on performance. *Journal of Business Economics* **90**:8, 1253-1284. [Crossref]

43. Andy Brownback, Sally Sadoff. 2020. Improving College Instruction through Incentives. *Journal of Political Economy* **128**:8, 2925-2972. [Crossref]

44. Robert Gillanders, Shadi Karazi, Fiona O'Riordan. 2020. Loss aversion as a motivator for engagement with peer assessment. *Innovations in Education and Teaching International* **57**:4, 424-433. [Crossref]

45. Jens Dietrichson, Trine Filges, Rasmus H. Klokker, Bjørn C. A. Viinholt, Martin Bøg, Ulla H. Jensen. 2020. Targeted school-based interventions for improving reading and mathematics for students with, or at risk of, academic difficulties in Grades 7–12: A systematic review. *Campbell Systematic Reviews* **16**:2. . [Crossref]

46. Bhagyashree Katare, Qihui Chen, Michael Wetzstein. 2020. Exam credits and exam performance in university economics courses – evidence from a regression discontinuity experiment. *Applied Economics Letters* **27**:9, 685-689. [Crossref]

47. Joshua Goodman, Oded Gurantz, Jonathan Smith. 2020. Take Two! SAT Retaking and College Enrollment Gaps. *American Economic Journal: Economic Policy* **12**:2, 115-158. [Abstract] [View PDF article] [PDF with links]

48. Shinya Kajitani, Keiichi Morimoto, Shiba Suzuki. 2020. Information feedback in relative grading: Evidence from a field experiment. *PLOS ONE* **15**:4, e0231548. [Crossref]

49. Tatiana Homonoff, Barton Willage, Alexander Willén. 2020. Rebates as incentives: The effects of a gym membership reimbursement program. *Journal of Health Economics* **70**, 102285. [Crossref]

50. Elisa De Marchi, Alessia Cavaliere, Rodolfo M. Nayga, Alessandro Banterle. 2020. Incentivizing Vegetable Consumption in School-Aged Children: Evidence from a Field Experiment. *Journal of Consumer Affairs* **54**:1, 261-285. [Crossref]

51. Vi-Nhuan Le. 2020. Do Student-Level Incentives Increase Student Achievement? A Review of the Effect of Monetary Incentives on Test Performance. *Teachers College Record: The Voice of Scholarship in Education* **122**:3, 1-34. [Crossref]

52. Gary Charness, Michael Cooper, J Lucas Reddinger. Wage Policies, Incentive Schemes, and Motivation 1-33. [Crossref]

53. Marie Claire Villeval. Performance Feedback and Peer Effects 1-38. [Crossref]

54. Bouke Klein Teeselink, Rogier J.D. Potter van Loon, Martijn J. van den Assem, Dennie van Dolder. 2020. Incentives, performance and choking in darts. *Journal of Economic Behavior & Organization* **169**, 38-52. [Crossref]

55. Manuela Angelucci, Silvia Prina, Heather Royer, Anya Samek. 2019. Incentives and Unintended Consequences: Spillover Effects in Food Choice. *American Economic Journal: Economic Policy* **11**:4, 66-95. [Abstract] [View PDF article] [PDF with links]

56. Uri Gneezy, John A. List, Jeffrey A. Livingston, Xiangdong Qin, Sally Sadoff, Yang Xu. 2019. Measuring Success in Education: The Role of Effort on the Test Itself. *American Economic Review: Insights* **1**:3, 291-308. [Abstract] [View PDF article] [PDF with links]

57. Christiane Bradler, Susanne Neckermann. 2019. The Magic of the Personal Touch: Field Experimental Evidence on Money and Appreciation as Gifts*. *The Scandinavian Journal of Economics* **121**:3, 1189-1221. [Crossref]

58. Damon Abraham, Kateri McRae, Jennifer A. Mangels. 2019. "A" for Effort: Rewarding Effortful Retrieval Attempts Improves Learning From General Knowledge Errors in Women. *Frontiers in Psychology* **10**. . [Crossref]

59. Hongyan Liu, Hao Xue, Yaojiang Shi, Scott Rozelle. 2019. The academic performance of primary school students from rural China. *China Agricultural Economic Review* **11**:2, 253-279. [Crossref]

60. David Hagmann, Emily H Ho, George Loewenstein. 2019. Nudging out support for a carbon tax. *Nature Climate Change* **9**:6, 484-489. [Crossref]

61. Gerhard Riener, Valentin Wagner. 2019. On the design of non-monetary incentives in schools. *Education Economics* **27**:3, 223-240. [Crossref]

62. Andy Brownback, Alex Imas, Michael Kuhn. 2019. Behavioral Food Subsidies. *SSRN Electronic Journal* . [Crossref]

63. Inga Deimen, Julia Wirtz. 2019. Control, Cost, and Confidence: Explaining Perseverance in the Face of Failure. *SSRN Electronic Journal* . [Crossref]

64. Tatiana A. Homonoff. 2018. Can Small Incentives Have Large Effects? The Impact of Taxes versus Bonuses on Disposable Bag Use. *American Economic Journal: Economic Policy* **10**:4, 177-210. [Abstract] [View PDF article] [PDF with links]

65. Oded Gurantz. 2018. A Little Can Go a Long Way: The Impact of Advertising Services on Program Take-Up. *Educational Evaluation and Policy Analysis* **40**:3, 382-398. [Crossref]

66. Felipe Balmaceda. 2018. Optimal task assignments with loss-averse agents. *European Economic Review* **105**, 1-26. [Crossref]

67. Hannah Schildberg-Hörisch, Philipp Thoste, Valentin Wagner. 2018. Nudging in der Schulmensa: Verhaltenswissenschaftliche Erkenntnisse können Ernährungsgewohnheiten positiv beeinflussen. *Vierteljahrshefte zur Wirtschaftsforschung* **87**:2, 109-125. [Crossref]

68. Jia Shuai, Shaocong Mo, Yunyao Yang, Xu Han, Zhongyuan Chen, Zefeng Xie, Chao Dai. College data analysis based on multi-learning method 5-8. [Crossref]

69. Andreas Ostermaier. 2018. Incentives for students: effects of certificates and deadlines on student performance. *Journal of Business Economics* **88**:1, 65-96. [Crossref]

70. Felipe Balmaceda. 2018. Optimal Task Assignments with Loss-Averse Agents. *SSRN Electronic Journal* . [Crossref]

71. Derek Lemoine. 2018. Rationally Misplaced Confidence. *SSRN Electronic Journal* **28**. . [Crossref]

72. Carly Robinson, Jana Gallus, Monica Lee, Todd Rogers. 2018. The Demotivating Effect (and Unintended Message) of Retrospective Awards. *SSRN Electronic Journal* . [Crossref]

73. Bouke Klein Teeselink, Rogier Potter van Loon, Martijn J. van den Assem, Dennie van Dolder. 2018. Incentives, Performance and Choking in Darts. *SSRN Electronic Journal* . [Crossref]

74. Catherine Insel, Erik K. Kastman, Catherine R. Glenn, Leah H. Somerville. 2017. Development of corticostriatal connectivity constrains goal-directed behavior during adolescence. *Nature Communications* **8**:1. . [Crossref]

75. . There is no learning without prepared, motivated learners 107-130. [Crossref]

76. . Overview: Learning to realize education's promise 1-35. [Crossref]

77. Alex Imas, Sally Sadoff, Anya Samek. 2017. Do People Anticipate Loss Aversion?. *Management Science* **63**:5, 1271-1284. [Crossref]