



Federal Reserve Bank of Chicago

**Financial Incentives and Educational
Investment: The Impact of
Performance-Based Scholarships
on Student Time Use**

Lisa Barrow and Cecilia Elena Rouse

August 2013

WP 2013-07

**Financial Incentives and Educational Investment:
The Impact of Performance-Based Scholarships on Student Time Use**

Lisa Barrow
lbarrow@frbchi.org
Federal Reserve Bank of Chicago

Cecilia Elena Rouse
rouse@princeton.edu
Princeton University

August, 2013

We thank Eric Auerbach, Laurien Gilbert, Ming Gu, Steve Mello, and Lauren Sartain for expert research assistance; Leslyn Hall and Lisa Markman Pithers with help developing the survey; and Elijah de la Campa and Reshma Patel for extensive help in understanding the MDRC data. Orley Ashenfelter, Alan Krueger, Jonas Fisher, Derek Neal, Reshma Patel, and Lashawn Richburg-Hayes as well as seminar participants at Cornell University, the Federal Reserve Bank of Chicago, Federal Reserve Bank of New York, Harvard University, Michigan State, Princeton University, University of Chicago, University of Pennsylvania, and University of Virginia provided helpful conversations and comments. Some of the data used in this paper are derived from data files made available by MDRC. We thank the Bill & Melinda Gates Foundation and the Princeton University Industrial Relation Section for generous funding. The authors remain solely responsible for how the data have been used or interpreted. Any views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System. Any errors are ours.

**Financial Incentives and Educational Investment:
The Impact of Performance-Based Scholarships on Student Time Use**

Abstract

Using survey data from a field experiment in the U.S., we test whether and how financial incentives change student behavior. We find that providing post-secondary scholarships with incentives to meet performance, enrollment, and/or attendance benchmarks induced students to devote more time to educational activities and to increase the quality of effort toward, and engagement with, their studies; students also allocated less time to other activities such as work and leisure. While the incentives did not generate impacts after eligibility had ended, they also did not decrease students' inherent interest or enjoyment in learning. Finally, we present evidence suggesting that students were motivated more by the incentives provided than simply the effect of giving additional money, and that students who were arguably less time-constrained were more responsive to the incentives as were those who were plausibly more myopic. Overall these results indicate that well-designed incentives can induce post-secondary students to increase investments in educational attainment.

I. Introduction

Educators have long been worried about relatively low levels of educational performance among U.S. students.¹ Elementary and secondary school students in the U.S. perform poorly on academic tests compared to their peers in other developed countries (Baldi et al., 2007), and at the post-secondary level there is increasing concern that while the U.S. has one of the highest rates of college attendance in the world, rates of college completion lag those of other countries (*OECD Indicators*, 2011). In response, educators have implemented many policies aimed at changing curricula, class size, teacher effectiveness, and other resources at all levels of schooling. More recently, there has been interest in targeting another key component: the effort exerted by students themselves towards their studies.

One approach to motivating students to work harder in school has been to offer them rewards for achieving prescribed benchmarks. Related to the conditional cash transfer strategies that have been growing in popularity in developing countries (see, e.g., Das et al., 2005 and Rawlings and Rubio, 2005), U.S. educators have implemented programs in which students are paid for achieving benchmarks such as a minimum grade-point average (GPA) or for reading a minimum number of books. These strategies are based on the belief that current pay-offs to education are too far in the future (and potentially too “diffuse”) to motivate students to work hard in school. As such, by implementing more immediate pay-offs, these incentive-based strategies are designed to provide students with a bigger incentive to work hard.

Unfortunately, the evidence to date of such efforts has yielded somewhat mixed, and often small, impacts on student achievement. For example, Jackson (2010a and 2010b) find some evidence that the Advanced Placement Incentive Program (APIP) in Texas—which

¹ For example, the 1983 report on American education, “A Nation at Risk” spurred a wave of concern regarding poor academic performance at nearly every level among U.S. students.

rewards high school students (and teachers) for AP courses and exam scores—increased scores on the SAT and ACT Tests, increased rates of college matriculation and persistence (students were more likely to remain in school beyond their first year of college), and improved post-secondary school grades. These results are similar to those reported by Angrist and Lavy (2009) from a high school incentive program in Israel. Somewhat in contrast, Fryer (2011) reports suggestive evidence that rewarding elementary and secondary school students for effort focused on education inputs, such as reading books, may increase test score achievement while rewarding them for education outcomes, such as grades and test scores, does not. At the post-secondary level, estimated impacts of incentives have also been modest. For example, Angrist, Lang, and Oreopoulos (2009) and Angrist, Oreopoulos, and Williams (2012) report small impacts at a four-year college in Canada on grades, although impacts may be larger for some subgroups.² Barrow, et al. (2012) report positive impacts on enrollment and credits earned for a program aimed at low-income adults attending community college in the U.S., and early results from MDRC's Performance-based Scholarship Demonstration (described below) suggest modest impacts on some academic outcomes (such as enrollment and credits earned) but little impact on the total number of semesters enrolled (see Richburg-Hayes et al., 2011; Cha and Patel, 2010; and Miller et al., 2011).

The fact that impacts of incentives on academic outcomes have been small, at best, raises the question of whether educational investments can be effectively influenced through the use of incentives in the sense that they actually change student behavior. Alternatively it is also possible that these small positive results arise from statistical anomaly, reflect the provision of additional income rather than the incentive structure, or represent changes along other

² Specifically, Angrist, Lang, and Oreopoulos (2009) report larger impacts for women, although the subgroup result is not replicated in Angrist, Oreopoulos, and Williams (2012). Angrist, Oreopoulos, and Williams (2012) report larger effects among those aware of the program rules.

dimensions such as taking easier classes. In this paper, we evaluate the effect of two performance-based scholarship programs for post-secondary students on a variety of outcomes, but especially on student effort as reflected in time use in order to understand whether incentives indeed induce students to increase effort toward their education. Students were randomly assigned to treatment and control groups where the treatments (the incentive payments) varied in length and magnitude and were tied to meeting performance, enrollment, and/or attendance benchmarks. To measure the impact of performance-based scholarships on student educational effort, we asked respondents about time use over the prior week and implemented a time diary survey. Further, variation in the incentive structure of the scholarships allows us to test a variety of hypotheses about the impacts of incentive payments. For example, we can test whether incentive payments are “habit forming” or otherwise change the achievement production function and examine whether incentive payments negatively impact “intrinsic” motivation, i.e., students’ inherent interest in and enjoyment of learning.

We find that students eligible for a performance-based scholarship (PBS) devoted more time to educational activities, increased the quality of effort toward, and engagement with, their studies, and allocated less time to other activities such as work and leisure. Additional evidence indicates that incentives did not affect student behavior—either positively or negatively—after incentive payments were removed and suggests that students were motivated by the incentives provided by the scholarships rather than simply the additional money. Finally, other analyses imply that students who were arguably less time-constrained were more responsive to the performance-based scholarships as were those who were plausibly more myopic. Overall our findings indicate that well-designed incentives can induce post-secondary students to increase investments in educational attainment. One remaining puzzle, however, is that larger incentive

payments did not seem to induce students to increase effort more than smaller incentive payments.

We next discuss a theoretical framework for thinking about effort devoted to schooling and the role of incentive scholarships (Section II). We describe the two interventions studied, the data, and sample characteristics of program participants in Section III. The estimation strategy and results are presented in Section IV, and Section V concludes.

II. Theoretical framework

We adopt the framework introduced by Becker (1967) in which students invest in their education until the marginal cost of doing so equals the marginal benefit. Assuming linearity for ease of exposition, suppose that student i 's GPA, g_i , depends on ability (and/or preparation) a_i , effort e_i , and some random noise ε_i as follows:

$$g_i = \delta_0 + \delta_1 e_i + \delta_2 a_i + \varepsilon_i, \quad (1)$$

where δ_0 , δ_1 , and δ_2 are all positive parameters of the GPA production function. Let ε be distributed $F(\varepsilon)$, with density $f(\varepsilon)$, and let $c(e)$ reflect the cost of effort. Assume $c(e)$ is an increasing, concave, and twice differentiable function. Further assume there is a payoff W for achieving a minimum GPA, g^{min} , with a payoff of zero otherwise.³ In our application, W can be thought of as the present discounted value of the earnings increase associated with additional college credits (net of tuition and other costs) plus incentive payments for eligible students. Basically, college only pays a return in terms of higher future earnings if one earns the college credits at a minimum level of proficiency (as reflected in GPA).

³ In fact, the payoff could be negative for students paying tuition but failing to pass any classes. One could also think about there being a payoff to each course completed with a minimum grade level, and higher payoffs to achieving grades above the minimum threshold.

Assuming students maximize utility by maximizing the net expected benefit of effort, the student's maximization problem is as follows:

$$\max_{e_i} \{ [1 - F(g^{\min} - \delta_0 - \delta_1 e_i - \delta_2 a_i)] \cdot W - c(e_i) \} \text{ s.t. } e_i \geq 0. \quad (2)$$

The optimal value of effort, e_i^* , is characterized by the following conditions:

$$\begin{aligned} c'(e_i^*) &\geq f(g^{\min} - \delta_0 - \delta_1 e_i^* - \delta_2 a_i) \cdot \delta_1 W, \\ e_i^* &\geq 0, \text{ and} \\ e_i^* [f(g^{\min} - \delta_0 - \delta_1 e_i^* - \delta_2 a_i) \cdot \delta_1 W - c'(e_i^*)] &= 0. \end{aligned} \quad (3)$$

If the marginal benefit is relatively low or the marginal costs are relatively high a student may not enroll or continue in college, setting $e_i^* = 0$. For a student who does enroll, an increase in the payoff to meeting the benchmark (a rise in W) will lead to an increase in effort toward her studies. Likewise a fall in W will lead to a decrease in effort.⁴

While we have written down a simple static model of optimal choice of effort, there are potential dynamic effects which we will consider and test in addition to potential roles for heterogeneity and possible unintended consequences. For example, traditional need-based and merit-based scholarships provide an incentive to enroll in college by effectively lowering the costs of enrolling in college regardless of whether the student passes her classes once she gets there. In the theoretical framework above, these types of scholarships have no impact on the marginal value or cost of effort conditional on enrolling in school. However, in a dynamic

⁴ Of course, a change in the payoff may also affect enrollment decisions for some students.

version of the model, future scholarship receipt may depend on meeting the g^{min} benchmark in the current semester. Pell Grants, for example, stipulate that future grant receipt depends on meeting satisfactory academic progress, but within-semester grant receipt is unaffected by performance. Performance-based scholarships, on the other hand, increase the marginal benefit of effort towards schoolwork in the current semester by increasing W . For example, payments may be contingent on meeting benchmark performance goals such as a minimum GPA. Because a PBS increases the short-run financial rewards to effort, we would expect PBS-eligible students to allocate more time to educationally-productive activities, such as studying, which should in turn translate into greater educational attainment, on average.

We may expect the effectiveness of PBS programs to depend on both the size of the scholarship as well as the impact of the program on a student's GPA production function. In the basic theoretical model above, the direct effect of increasing the size of the incentive payment in the current semester is a contemporaneous increase in effort but no impact on effort in future semesters. However, there may also be indirect effects leading to impacts in future semesters. For example, an increase in effort in the current semester may reduce the marginal cost of effort in the future (suppose increased studying today is "habit-forming"), leading PBS eligibility to have a positive impact on student effort after eligibility has expired.⁵ Similarly, if increased effort today teaches students how to study more effectively (by increasing the marginal benefit of effort through an increase in δ_1), then PBS eligibility could also have a lasting positive impact by increasing effort after eligibility has expired.

At the same time, cognitive psychologists worry that while incentive payments may motivate students to do better in the short term, the students may be motivated for the "wrong"

⁵ This is similar to the behaviorist theory in psychology described by Gneezy and Rustichini (2000) that suggests incentive payments tied to studying (which requires effort) will lead to a positive (or at least less negative) association with studying in the future.

reasons. They distinguish between internal (or intrinsic) motivation in which a student is motivated to work hard because he or she finds hard work inherently enjoyable or interesting and external (or extrinsic) motivation in which a student is motivated to work because it leads to a separable outcome (such as a performance-based scholarship incentive payment) (see, e.g., Deci (1975) and Deci and Ryan (1985)). A literature in psychology documents more positive educational outcomes the greater the level of “internalization” of the motivation (e.g., Pintrich and De Groot, 1990). As such, one potential concern regarding performance-based rewards in education is that while such scholarships may increase external motivation, they may decrease internal motivation (e.g., Deci, Koestner, and Ryan, 1999). In the model above, a reduction in intrinsic motivation could be viewed as raising students’ cost of effort. Assuming this effect is permanent, we would expect to see a negative impact of PBS eligibility on effort in future semesters after PBS eligibility has expired (Huffman and Bognanno, 2012; and Benabou and Tirole, 2008). Some even hypothesize that the reduction in intrinsic motivation may more than offset the external motivation provided by the incentive. This would lead to a negative impact of PBS eligibility even during semesters in which incentive payments were provided (Gneezy and Rustichini, 2000).

There may also be some heterogeneity in responsiveness across students. Whether and by how much a performance incentive changes an individual student’s effort will depend on ability and the marginal cost of effort. Suppose the density $f()$ in equation (3) is roughly normally distributed with small values in the tails. For a high-ability student, increasing the payoff W will have little effect on her effort because she will essentially be able to meet the minimum GPA requirement on ability with no increased effort. The reason is her prior ability alone made it likely she would achieve the GPA benchmark. For students with lower ability or who are less

well prepared, the performance incentive will induce them to increase effort in order to increase the probability of meeting the minimum GPA requirement.⁶ We also expect that the performance incentive will have bigger impacts on effort for students who heavily discount future benefits as the returns to schooling in terms of future earnings should already motivate less-myopic students but may be too far in the future to have an impact on more present-oriented students; effectively, myopic students perceive a lower W . On the cost side (all else equal), there may be heterogeneity in cost functions, and we would expect to see students facing a higher marginal cost of effort (such as those with young children) to have a smaller change in effort in response to changes in the payoff W than students facing a lower marginal cost of effort.

Finally, while the intention of a performance-based scholarship is to increase student effort in educationally productive ways, it may unintentionally increase student attempts to raise their performance in ways that are not educationally productive. For example, Cornwell, Lee, and Mustard (2005) find that the Georgia HOPE scholarship which had grade incentives but not credit incentives reduced the likelihood that students registered for a full credit load and increased the likelihood that students withdrew from courses presumably to increase the probability that they met the minimum GPA benchmark. Other unintended consequences could include cheating on exams, asking professors to regrade tests and/or papers, or taking easier classes.

III. The Scholarship Programs and the Time-Use Survey

The data we analyze were collected as part of the Performance-Based Scholarship Demonstration conducted by MDRC at 8 institutions and at a state-wide organization in

⁶ If the benchmark is set too high, however, students with a low probability of meeting the benchmark are unlikely to increase effort because the probability of meeting the GPA benchmark, even with high levels of effort, is too low and effort is costly.

California. The structure of the scholarship programs as well as the populations being studied vary across the sites; this study presents results from a supplementary “Time Use” module we implemented at sites in New York City (NYC) and California (CA).⁷ In both demonstrations analyzed, the scholarships supplemented any other financial aid for which the students qualified (such as federal Pell Grants and state aid), although institutions may have adjusted aid awards in response to scholarship eligibility for some students.⁸

A. The Scholarship Programs

The New York City Program

The intervention in New York City was implemented at two campuses of the City University of New York System (CUNY) – the Borough of Manhattan Community College (BMCC) and Hostos. Students were recruited on campus in three cohorts (Fall 2008, Spring 2009, and Fall 2009). Eligible students were aged 22-35, had tested into (and not yet passed) at least one developmental course, were eligible for a federal Pell Grant, enrolled in at least 6 credit or contact hours (at the time of “intake”), and lived away from their parents. Once program staff determined eligibility for the study, students who agreed to participate provided baseline demographic information and were randomly assigned by MDRC to the program or control groups. Everyone who attended an orientation session (at which they were introduced to the study) and signed up to participate in the study received a \$25 metro card.

⁷ See Richburg-Hayes (2009) for more details on the programs in each site in the larger demonstration.

⁸ Institutions are required to reduce aid awards when the financial aid award plus the outside scholarship exceeds financial need by more than \$300. In practice, institutions generally treat outside scholarship earnings favorably by reducing students’ financial aid in the form of loans or work study or applying the scholarship to unmet financial need. (See www.finaid.org/scholarships/outside.phtml.) Other MDRC studies have found that PBS-eligible participants received net increases in aid and/or reductions in student loans relative to their control group. See Cha and Patel (2010), Miller et al. (2011), and Patel and Rudd (2012).

Students randomly assigned to a program (treatment) group were eligible to receive a performance-based scholarship worth up to \$1,300 each semester for two semesters (for a total of \$2,600).⁹ As the goal of this scholarship was to reward attendance (an input to academic success) as well as performance at the end of the semester, the incentive payments were structured as follows: After registering for at least six credits (meaning tuition had been paid or a payment plan had been established) the student received \$200; with “continued enrollment at mid-semester” he or she received \$450; and with a final grade of “C” or better (or having passed developmental education) in at least 6 credits (or equated credits), he or she received \$650.¹⁰ So as not to discourage students mid-semester, if a student missed the mid-semester payment it could be recouped at the end of the semester if the final requirement was met. The students were also eligible for the entire incentive a second semester, independent of having met any of the first semester benchmarks.¹¹

The California Program

The California program is unique in the PBS demonstration in that random assignment took place in the spring of the participants’ senior year of high school and students could use the scholarship at any accredited institution.¹² Individuals in the study were selected from participants in “Cash for College” workshops at which attendees were given assistance in

⁹ In reality in NYC some students were randomly assigned to a second treatment group that was eligible to receive the performance-based scholarship during the regular semesters plus a performance-based scholarship for one consecutive summer worth up to \$1,300 (for a total of \$3,900). Because we focus on regular semester outcomes during which the incentive structures for the two treatment groups are identical, we do not present results separately for the two treatment groups.

¹⁰ “Continued enrollment at mid-semester” was determined by whether the student attended class at least once in the first three weeks of the semester and at least once during the fourth or fifth weeks of the semester. Equated credits are given in developmental education classes and do not count towards a degree or certificate.

¹¹ See Richburg-Hayes, Sommo, and Welbeck (2011) for more background on the New York demonstration.

¹² At the other sites, the scholarships were tied to enrollment at the institution at which the student was initially randomly assigned. In addition, all of the other study participants were at least “on campus” to learn about the demonstration suggesting a relatively high level of interest in, and commitment to, attending college. See Ware and Patel (2012) for more background on the California program.

completing the Free Application for Federal Student Aid (FAFSA). In order to be eligible for the study, they also had to complete the FAFSA by March 2 of the year in question. Study participants were selected from sites in the Los Angeles and Far North regions in 2009 and 2010 and from the Kern County and Capitol regions in 2010. Randomization occurred within each workshop in each year. Because the students were high school seniors at the time of random assignment, this demonstration allows us to determine the impact of these scholarships not only on persistence among college students, but also on initial college enrollment.

To be eligible for this study, participants had to have attended a Cash for College workshop in one of the participating regions; been a high school senior at the time of the workshop; submitted a Free Application for Federal Student Aid (FAFSA) and Cal Grant GPA Verification Form by the Cal Grant deadline (early March); met the low-income eligibility standards based on the Cal Grant income thresholds; and signed an informed consent form or had a parent provide consent for participation.

The incentive varied in length (as short as one semester and as long as four semesters), size of scholarship (as little as \$1,000 and as much as \$4,000), and whether there was a performance requirement attached to it. Also, the performance-based scholarships were paid directly to the students whereas the non-PBS was paid to the institution. Table 1, below, shows the structure of the demonstration for the Fall 2009 cohort more specifically; the structure was similar for the Fall 2010 cohort. There were six treatment groups labeled 1 to 6 in Table 1. Group 1 was randomly selected to receive a California Cash for College scholarship worth \$1,000 which is a typical grant that has no performance component and is paid directly to the student's institution. Groups 2 through 6 had a performance-based component with payments made directly to the student. Group 2 was randomly assigned to receive \$1,000 over one

academic term (a semester or quarter); group 3 was randomly selected to receive \$500 per semester for two semesters (or \$333 per quarter over three quarters); group 4 was selected to receive \$1,000 per semester for two semesters (or \$667 per quarter over three quarters); group 5 was to receive \$500 per semester for four semesters (or \$333 per quarter over six quarters); and group 6 was to receive \$1,000 per semester for four semesters (or \$667 per quarter over six quarters). During fall semesters of eligibility, one-half of the PBS was paid conditional on enrolling for six or more credits at an accredited, degree-granting institution in the U.S., and one-half was paid if the student met the end-of-semester benchmark (a final average grade of “C” or better in at least 6 credits). During spring semesters of eligibility, the entire scholarship payment was based on meeting the end-of-semester benchmark.

Table 1: Structure of the California Program

Scholarship Type	Total Amount	Performance Based?	Duration	Fall 2009		Spring 2010	Fall 2010		Spring 2011
				Initial	Final		Initial	Final	
1	\$1,000	No	1 term	\$1,000					
2	\$1,000	Yes	1 term	\$500	\$500				
3	\$1,000	Yes	1 year	\$250	\$250	\$500			
4	\$2,000	Yes	1 year	\$500	\$500	\$1,000			
5	\$2,000	Yes	2 years	\$250	\$250	\$500	\$250	\$250	\$500
6	\$4,000	Yes	2 years	\$500	\$500	\$1,000	\$500	\$500	\$1,000

Source: Ware and Patel (2012). The dates refer to the incentive payouts for the 2009 cohort but the structure is the same for the 2010 cohort. The schedule shown applies to institutions organized around semesters; for institutions organized into quarters the scholarship amount is the same in total but the payments are divided into three quarters in the academic year.

In addition, aside from the institution being accredited, there were no restrictions on *where* the participants enrolled in college. That said, according to data from the Cash for College workshops, among the two-thirds of students who enroll in college the following fall over 90%

attend a public college or university within California, about one-half in a two-year college and the other half in a four-year institution.

B. Numbers of Participants

In Table 2 we present information on the number of students in each cohort in each demonstration. In total 6,662 individuals were recruited to be part of the PBS study; 2,474 were randomly assigned to the program-eligible group and 4,188 were assigned to the control group.¹³ We also surveyed an additional 1,500 individuals in California as part of the control group; they were randomly selected from non-study group individuals who were not selected to be in either the program or control group for the MDRC study. Appendix Tables A1a and A1b show means of background characteristics (at baseline) by treatment/control status. While there are one or two characteristics that appear to differ between treatment and control groups, an omnibus f-test yielded a p-value of 0.68 in Appendix Table A1a for NYC and 0.48 in Appendix Table A1b for CA suggesting that randomization successfully balanced the two groups, on average.

According to Richburg-Hayes and Patel (forthcoming) nearly all (99%) of the treatment students in NYC received the initial payment the first semester and 97% received the midterm payment (that required continued enrollment, as defined above); 72% received the performance-based payment at the end of the term. Scholarship receipt was lower in the second semester with only 83% receiving the first payment, 80% receiving the second, and 58% receiving the final payment. In CA, initial results for cohort 1 reported in Ware and Patel (2012) indicate that 85 percent of the scholarship eligible participants received an enrollment payment and of those, 60 percent earned the performance-based payment at the end of the fall 2009 semester.

¹³ We did not receive contact information for one individual in New York so the total number of individuals we attempted to survey in New York and California combined is 8,161.

C. Time-Use Survey¹⁴

To better understand the impact of performance-based scholarships on student educational effort, we implemented an independent (web-based) survey of participants. We asked respondents general questions about educational attainment and work (roughly based on the *Current Population Survey*). The centerpiece included two types of questions designed to better understand how respondents allocated their time to different activities. To understand more “granular” time allocation we implemented a time diary for which we used the *American Time Use Survey* (ATUS) as a template. Accounting for an entire 24-hour time period, the ATUS asks the respondent to list his or her activities, describe where the activities took place, and with whom. In addition we included questions about time use over the last 7 days to accommodate those activities that are particularly relevant to students and for which it would be valuable to measure over longer periods (such as time spent studying per day over the past week). Participants were offered an incentive to participate.

In addition to the questions regarding time use over the previous 24 hours or 7 days (that reflect the “quantity” of time allocated to activities), the survey also included questions to measure the quality of educational efforts. To capture these other dimensions of effort, we included questions on learning strategies, academic self-efficacy, and motivation. To measure learning strategies that should help students perform better in class, we included questions from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al. 1991). The scale consists of five questions on a seven-point scale with questions such as: “When I become confused about something I’m reading, I go back and try to figure it out” (responses range from

¹⁴ We only briefly describe the survey in this section. See Barrow and Rouse (2013) for more details on the survey design and implementation.

not at all true (1) to very true (7)). In addition, researchers have documented a link between perceived self-efficacy (e.g., an individual's expectations regarding success or assessment of his or her ability to master material) and academic performance (see, e.g., Pintrich and De Groot 1990). Therefore we included five questions that form a scale to capture perceived academic efficacy (the *Patterns of Adaptive Learning Scales* (PALS) by Midgley et al. 2000). These questions are of the form, "I'm certain I can master the skills taught in this class this year." with responses on a similar seven-point scale.

Finally, we attempted to assess whether the incentives also induced unintended consequences such as cheating, taking easier classes, or attending classes simply to receive the reward not because of an inherent interest in the academics (which some psychologists argue can ultimately adversely affect academic achievement, as discussed earlier). As such, on the survey we asked participants about life satisfaction, whether they had taken "challenging classes," if they had ever asked for a regrade, and if they had ever felt it necessary to cheat. To capture external and internal motivation, we asked questions of both current students and those not currently enrolled along the lines of, "If I do my class assignments, it's because I would feel guilty if I did not" (also on a seven-point scale).¹⁵

We focus this analysis on time use and effort in the first semester after random assignment as the majority of both program and control students were enrolled in a post-secondary institution such that an analysis of time use is most compelling as one of the factors that determine educational success.¹⁶ That said, we also surveyed each cohort in the second

¹⁵ Specifically, "external motivation" is the mean of two questions: "If I attend class regularly it's because I want to get a good grade" and "If I raise my hand in class it's because I want to receive a good participation grade." "Internal motivation" is the mean two questions: "If I turn in a class assignment on time it's because it makes me happy to be on time" and "If I attend class often it's because I enjoy learning."

¹⁶ Because we focus on the first semester after random assignment, we do not include data from the first cohort in New York City as we were only able to first survey them in the second semester after random assignment.

semester after random assignment to gauge the extent to which time use changed and whether differences between the program and control group members persisted. Overall we achieved an average response rate of about 73% in New York City and about 57% in California in terms of the percentage of participants who ever responded to a survey.

Table 3 presents selected mean baseline characteristics for study participants at the time of random assignment and compares them to nationally-representative samples of students from the *National Postsecondary Student Aid Study (NPSAS) of 2008* designed to be comparable to the participants in the study.¹⁷ In both sites there were slightly more women than men. Further, the proportion of Hispanics and blacks was much higher in the study sites than nationally. For example, in NYC 80% of the participants were black or Hispanic compared to 40% nationally; in CA 63% of participants were Hispanic compared to 15% nationally, and only 4% were black. Similarly, in both sites a language other than English is likely to be spoken. That said, aside from the racial/ethnic composition, the baseline characteristics of the study participants resemble those of other post-secondary students nationally.

IV. Estimation and Results

A. Empirical Approach and Sample

Below we present estimates of the effect of program eligibility on a variety of outcomes.

We model each outcome Y for individual i as follows:

$$Y_i = \alpha + \beta T_i + \mathbf{X}_i \Theta + \mathbf{p}_i \gamma + \nu_i, \quad (4)$$

¹⁷ The baseline data were collected by MDRC at the time participants were enrolled in the study and before they were randomly assigned to a program, control, or non-study group. The samples from the NPSAS have the same age ranges as the two sites. See the notes to the tables for other sample restrictions.

where T_i is a treatment status indicator for individual i being eligible for a program scholarship, \mathbf{X}_i is a vector of baseline characteristics (which may or may not be included), \mathbf{p}_i is a vector of indicators for the student's randomization pool, v_i is the error term, and α , β , Θ , and γ are parameters to be estimated; β represents the average effect on outcome Y of being randomly assigned to be eligible for the scholarship. In some specifications, we allow for a vector of treatment indicators depending on the type of scholarship for which the individual was eligible.

To facilitate interpretation and to improve statistical power, we group impacts on individual time use into two “domains” of most interest for this study: academic activities and non-academic activities.¹⁸ Further, we also summarize impacts of measures that reflect the quality of educational effort and those that capture potential “unintended consequences.”

To see how we analyze the effect of eligibility to receive a PBS on a “domain,” we note that we can rewrite equation (4) to obtain an effect of the treatment on each individual outcome, where k refers to the k th outcome:

$$Y_k = \alpha_k + \beta_k T + \mathbf{X}\Theta_k + \mathbf{p}\gamma_k + v_k = \mathbf{A}\Phi_k + v_k. \quad (5)$$

We can then summarize the individual estimates using a seemingly-unrelated regression (SUR) approach (Kling and Liebman, 2004). This approach is similar to simply averaging the estimated effect of being randomly assigned to be eligible for a PBS, if there are no missing values and no covariates.

¹⁸ As an alternative mechanism of grouping outcomes we conducted a factor analysis to identify empirically-determined principal components. The results roughly suggested that variables reflecting academic effort should be grouped together and those reflecting time spent on non-academic time should be grouped together. That said, we prefer our approach because it is more intuitive and it is possible to identify exactly which outcomes contribute to each domain.

More specifically, we first estimate equation (5) (or variants) and obtain an item-by-item estimate of β (i.e., β_k). We then standardize the estimates of β_k by the standard deviation of the outcome using the responses from the control group of participants (σ_k). The estimate of the impact of eligibility on time use and individual behavior is then the average of the standardized β s within each domain,

$$\beta_{AVG} = \frac{1}{K} \sum_{k=1}^K \beta_k / \sigma_k. \quad (6)$$

We estimate the standard errors for β_{AVG} using the following seemingly-unrelated regression system that allows us to account for the covariance between the estimates of β_k within each domain:

$$Y = (I_K \otimes A)\Phi + \mathbf{v} \quad Y = (Y'_1, \dots, Y'_K)' \quad (7)$$

where I_K is a K by K identity matrix and A is defined as in equation (5). We calculate the standard error of the resulting summary measure as the square root of the weighted sum of the variances and covariances among the individual effect estimates. One potential advantage of the SUR is that while estimates of each β_k may be statistically insignificant, the estimate of β_{AVG} may be statistically significant due to covariation among the outcomes. We present estimates of the original underlying regressions as well as those using the summary measure (i.e. the outcomes grouped together within a domain).

From our data, we focus on respondents to the first survey administered in the first semester after random assignment although we also report some results based on second

semester surveys. After dropping individuals who did not complete the time diary (or who had more than four “non-categorized” hours in the 24-hour time period) and those for whom we did not have data in the first part of the survey (due to an error by the survey contractor), we have data from 2,874 complete surveys in CA and 613 surveys in NYC. These complete surveys represent 92% and 93% (respectively) of the total number of survey respondents. Appendix Tables A2a and A2b show means of background characteristics (at baseline) by treatment/control status for our analysis sample. Again, while there are one or two characteristics that appear to differ between treatment and control groups, omnibus f-tests suggest that the two groups remain balanced, on average.

B. Basic Program Impacts on Educational and Other Outcomes

In Table 4a we present estimates of the effect of program eligibility on individual measures of time use based on our survey; in this table we do not distinguish between the types of performance-based scholarships offered in each of the two sites. In column (1) we provide outcome means for the control group participants in New York City, and in column (4) we provide outcome means for the control group participants in California. Program effect estimates with standard errors in parentheses are presented in column (2) for New York and columns (5) and (6) for California. Note that the estimates in column (5) reflect the impact of being eligible for a PBS while the estimates in column (6) reflect the impact of being eligible for a non-PBS. The p-value corresponding to the test that the PBS program impact equals the non-PBS program impact is presented in column (7). Program effects are estimated including controls for “randomization pool” fixed effects but no other baseline characteristics.¹⁹

¹⁹ For NYC, randomization pool fixed effects reflect the community college and cohort in which the participant was recruited. For CA, these fixed effects reflect the workshop region—Los Angeles, Far North, Kern County, or

In New York City we find that program-eligible students are no more likely to report ever enrolling in a post-secondary institution since random assignment than those in the control group. This result is not particularly surprising given that students were on campus when they were recruited for the program and needed to have registered for at least 6 credits (or equated credits) in order to be eligible to participate in the study. As evidence, 92% of NYC control-group students report ever attending a post-secondary institution since random assignment. In contrast there are larger differences in outcomes reflecting student effort. For example, the results suggest that eligibility for a performance-based scholarship induced participants to devote about 30 more minutes to educational activities in the prior 24-hour period than those assigned to the control group, although the difference is not statistically significant. And while control group students report spending about 2.8 hours per day studying in the last seven days, the PBS-eligible students report devoting 8 percent (or 13 minutes) more time to studying per day, although the difference is not statistically significant at conventional levels. Further, 78 percent of control group students report having attended most or all of their classes in the last seven days compared with 84 percent of students eligible for a performance-based scholarship, a difference that is statistically significant at the 10 percent level.

Results in the remaining columns of the table are from the California demonstration. Recall that a key difference between the CA program and that in NYC is the students in CA were randomly assigned in the spring of their senior year in high school while we surveyed them in the fall of what would be their first year attending a post-secondary institution. Focusing on the coefficients reported in column (5) of the table, we find that PBS-eligible students were 5.2 percentage points more likely than the control group to report ever enrolling at a post-secondary

Capital—and cohort in which the participant was recruited. Estimates are similar if we control for baseline characteristics such as age, sex, race, and parental education.

institution, a difference that is statistically significant at the 1 percent level. Further, the PBS-eligible students reported studying about 9 minutes more per day than those in the control group, were 7.3 percentage points more likely to have been prepared for class in the last 7 days, and were 6.7 percentage points more likely to report attending all or most of their classes in the last 7 days.

An important dimension to the demonstration in CA was the inclusion of a treatment group that was eligible for a “regular” scholarship that did not require meeting performance benchmarks. In particular, as discussed earlier, this non-PBS does not affect the marginal value of effort because payment is not tied to meeting benchmarks and is only valid for one semester. We generally find that the impacts are larger for those eligible for a PBS than for those offered a non-PBS; however, in most cases we are unable to detect a statistically significant difference. Tests of the difference in impact between the PBS and the non-PBS also potentially provide insight into whether students are responding to the incentives in the PBS or the additional income. We discuss this implication, below.

Before turning to how participants allocated their time to other activities, we consider two measures that may indicate ways of increasing academic effort without necessarily spending more time studying, namely, learning strategies and academic self-efficacy. As discussed above, PBS eligibility may induce participants to concentrate more on their studies by encouraging them to employ more effective study strategies making the time devoted to educational activities more productive. Similarly, by raising their academic self-efficacy the scholarships may also induce students to be more engaged with their studies. Results using scales based on the MSLQ Learning Strategies index and the PALS academic self-efficacy index are presented in the last two rows of Table 4a. We have standardized the variables using their respective control group

means and standard deviations; the coefficients therefore reflect impacts in standard deviation units. For both NYC and CA we estimate that eligibility for a PBS had positive and statistically significant impacts on these dimensions that range from 14 to 23 percent of a standard deviation. Note as well, the impacts on learning strategies and academic self-efficacy for those in CA selected for a non-PBS were substantially smaller than those selected for a PBS, consistent with increased academic effort on the part of PBS-eligible individuals.

Results presented thus far generally suggest that participants selected for a PBS devoted more time and effort to educational activities. Given there are only 24 hours in the day, a key question is what did PBS-eligible participants spend less time doing? Table 4b presents results from three other broad time categories based on the 24-hour time diary: work, household production, and leisure and other activities.²⁰ In NYC we estimate that the typical participant (as represented by the control group) works about 2.5 hours per day, devotes nearly 12 hours to home production (which includes sleeping), and devotes about 5 hours to “leisure.” We find that PBS-eligible participants accommodated spending about 30 more minutes in the last 24 hours on educational activities by devoting about 41 fewer minutes to leisure activities, an impact that is statistically significant at the 5 percent level. The typical participant in CA spends less time working and correspondingly more time in leisure activities than the NYC participants. However, we still find that participants in CA accommodated increased time spent on educational activities by spending (statistically) significantly less time on leisure activities, including reducing the number of nights out for fun during the past week. We find no evidence that PBS (or non-PBS) eligibility induced participants to reduce time spent on work or home

²⁰ “Home production” includes time spent on personal care, sleeping, eating and drinking, performing household tasks, and caring for others. “Leisure activities” include participating in a cultural activity, watching TV/movies/listening to music, using the computer, spending time with friends, sports, talking on the phone, other leisure, volunteering, and religious activities.

production; for both sites the estimated PBS impacts are small, (mostly) positive, and not statistically different from zero.

Finally, concerns about using incentives for academic achievement include the possibility of unintended consequences of the programs, such as cheating or taking easier classes to get good grades, or reducing students' internal motivation to pursue more education. In the bottom rows of Table 4b we present impacts on several potential unintended consequences for the participants in both the NYC and CA sites. In NYC we find little systematic evidence that eligibility for a PBS resulted in adverse outcomes. For example, those randomly selected for a PBS were more likely to report being satisfied with life and having taken challenging classes and were less likely to report having asked for a re-grade or having felt they had to cheat (only the impact on life satisfaction is statistically significant (at the 10 percent level)). Further, they were significantly more likely to report behavior consistent with increased internal motivation. In other words, the incentive payments did not seem to reduce their internal motivation.

In contrast, the results regarding unintended consequences are more mixed in CA. For example, on the one hand those in CA who were eligible for a PBS were more satisfied with life and more likely to take challenging classes compared to the control group (a difference that is statistically significant at the 1 percent level). On the other hand, PBS-eligible participants reported an increase in behavior that is consistent with external motivation compared to both control group participants and those randomly selected for a non-PBS.

Overall, the results in Tables 4a and 4b suggest that eligibility for a scholarship that requires achieving benchmarks results in an increase in time and effort devoted to educational activities with a decrease in time devoted to leisure. Further, there is at best mixed evidence that the same incentives result in adverse outcomes, such as cheating, "grade grubbing," or taking

easier classes. However, for many of the outcomes the impacts are not statistically different from zero.

To improve precision, in Table 5 we combine the individual outcomes into four “domains” using the SUR approach described above. Specifically we focus on academic activities, quality of educational input, non-academic activities, and unintended consequences.²¹ The impacts reported in Tables 5 and subsequent tables have been standardized such that they represent average impacts as a percentage of the control group standard deviation. Note that now we estimate a positive impact on academic activities of about 10 percent of a standard deviation in both NYC and CA and that both impacts are statistically significant at the 5 percent level. We also continue to estimate a positive and statistically significant impact on the quality of educational effort. In addition, we estimate a reduction in non-academic activities although the coefficient estimate is not significantly different from zero in NYC and is only significant at the 10 percent level in CA. Further, in both NYC and CA we estimate that, overall, there is not an increase in “unintended consequences” as a result of the academic financial incentive. In sum, these results suggest that scholarship incentives change student time allocation in the sense that they spend more time and effort on academic activities and less time on other activities.²²

²¹ “Educational activities” includes: “Hours spent on all academics in the last 24 hours,” “Hours studied in past 7 days,” “Prepared for last class in last 7 days,” and “Attended most/all classes in last 7 days.” “Quality of educational input” includes “Academic self-efficacy” and “MSLQ index.” “Non-academic activities” includes “Hours on household production,” “Hours on leisure,” “Nights out for fun in the past 7 days,” “Hours worked in last 24 hours,” and “Hours worked in the past 7 days.” And “Unintended consequences” includes “Strongly agree/agree have taken challenging classes,” “Ever felt had to cheat,” “indices of external motivation and internal motivation,” “Ever asked for a re-grade,” and “Very satisfied/satisfied with life.” We do not include whether an individual had “ever enrolled” in a post-secondary institution in the “all academic activities” index as it represents an academic decision on the extensive margin rather than the intensive margin, and NYC participants were recruited on campus after they had made the decision to enroll.

²² To explore the possibility that results (particularly in CA) are driven by an incentive effect only on the extensive margin, we have re-estimated the impacts limiting the samples to those students who enrolled in school. We find that estimated impacts on academic activities are somewhat smaller but still statistically different from zero suggesting that PBS eligibility affects both the extensive and intensive margins. Of course this relies on the assumption that those who are induced by the scholarship to enroll in school would not have otherwise put forth more effort toward their studies.

C. Impacts by Size, Duration, and Incentive Structure of the Scholarship

Three key questions are whether the size of the potential scholarship affects the impact on student behavior, whether scholarship eligibility impacts student behavior even after incentives are removed, and whether it is the incentive structure or the additional income that generates changes in student behavior. Prior studies of performance-based scholarships have tended to focus on one type of scholarship of a particular duration with variation in other types of resources available to students (such as student support services or a counselor). In CA students eligible for a performance-based scholarship were also randomly assigned to scholarships of differing durations and/or sizes as well as a non-performance-based scholarship. As noted in Table 1, CA students selected for scholarship eligibility were assigned to one non-incentive scholarship worth \$1,000 for one term or to one of five types of incentive scholarships that ranged from \$1,000 for one term to \$1,000 for each of four terms or \$500 for each of two terms to \$500 for each of four terms. We exploit this aspect of the design of the demonstration in CA to study the impact of scholarship characteristics on student behavior.²³ We present results using SUR in Table 6.²⁴

Impacts by Size of the Scholarship

Theoretically one would expect that the larger-sized scholarships would induce increased effort compared to smaller-sized scholarships during the semesters for which the students were eligible. As such, in the first semester after random assignment, we might expect to see a

²³ While we are able to test for some dimensions over which we would expect to see impacts by the characteristics of the scholarship, we also note that we only followed the students for at most two semesters after random assignment and therefore cannot test all dimensions on which the scholarship structure might matter.

²⁴ Estimates of program impacts for each of the underlying outcomes presented in this and the subsequent tables are available from the authors on request.

difference between scholarships worth \$1,000 and those worth \$500 per term. In Panel A of Table 6 we begin by examining whether larger scholarships generated larger impacts than smaller scholarships. Using results from the first academic term after random assignment (fall), impact estimates of the \$500/term scholarships are presented in column (1) and the \$1,000/term scholarships in column (2). P-values for the test of equality of the coefficient estimates in columns (1) and (2) are presented in column (4). Interestingly, we do not find large differences in the effect of PBS eligibility related to the size of the scholarship. Students who were eligible for a \$500 per semester scholarship responded similarly on most outcomes to students who were eligible for a \$1,000 per semester scholarship suggesting that larger incentive payment amounts did not lead to larger impacts on student effort. While this result is familiar in the context of survey implementation where experimental evidence suggests that larger incentives do not increase response rates (see, e.g., James and Bolstein (1992)), other laboratory and field experiments have found that paying a larger incentive improves performance relative to a smaller incentive (e.g., Gneezy and Rustichini (2000) and Lau (2013)). This finding remains a puzzle, although we offer some potential explanations for further consideration in the conclusion.

Impacts by Duration of the Scholarship

If incentives have no effect on the GPA production function (i.e. δ_1 and δ_2 in equations (2) and (3) are fixed), then one would expect that only the longer-duration scholarships would affect effort during the additional semesters of eligibility. As we note above, however, some of the literature on incentives and motivation predicts that we might observe reductions in effort after incentives are removed if the PBS has a negative effect on intrinsic motivation (Huffman and Bognanno, 2012; Benabou and Tirole, 2003). Alternatively, one might expect that increased

effort in the first semester might be habit-forming or make students more efficient at transforming study time into GPA in the second semester. In the latter case, PBSs may continue to have positive impacts on student outcomes in semesters after eligibility has expired. In Panel B of Table 6, we look at results for outcomes measured in the second semester after random assignment and consider impacts for the one-term scholarships that have expired (PBS and non-PBS) and the four PBSs for which eligibility continued two or more terms.

To begin, the results reported in column (1) examine the impacts of PBS eligibility on second semester student outcomes for participants who are no longer eligible for PBS payments. We find that the impacts of PBSs are largely contemporaneous. We find no difference in enrollment probabilities or the index of all academic activities between one-term PBS eligible participants and the control group during the second program semester. However, there is suggestive evidence of a lasting positive impact on the quality of educational inputs. Namely, one-term PBS eligible students have higher quality of effort than the control group in the semester after eligibility has expired, but the result is only statistically significant at the 10 percent level.

The results in column (2) represent the impacts of PBS eligibility in the second semester after random assignment for those who continue to be eligible for PBS payments. Here we continue to find positive impacts of PBS eligibility on academic effort and quality of effort relative to control group participants, and we find negative impacts of PBS eligibility on non-academic activities and unintended consequences. As such, we find that a performance-based scholarship primarily affects student behavior in the semester in which the student is eligible for the scholarship. In contrast to predictions from some dynamic models, we do not detect a negative impact of incentive eligibility on educationally-productive behavior once the incentive

is removed nor do we find strong evidence of a lasting change in student behavior as a result of prior eligibility for an incentive.

Incentives versus Additional Income

Finally an important question regarding any results with performance-based scholarships is whether the impacts are driven by the additional income or by the incentive structure of the scholarship. In other words, does it matter that the PBS comes with an incentive structure or would a simple monetary award with no incentives generate the same impacts? In our study, a comparison of the (fall term) impacts of the non-PBS (worth \$1,000 in one term) and the PBS of \$1,000 per term potentially provides a test of the impact of the incentive structure in the performance-based scholarship compared with just awarding additional money. This test can be made in Panel A of Table 6 by comparing the impacts of the \$1,000 PBSs (column 2) to those of the \$1,000 per term non-PBS in the first term (column 3) (the p-values of the tests of equality are in column (5)). With the exception of the coefficient on “unintended consequences,” we find that the magnitudes of the PBS coefficient estimates are larger in absolute value than those for the non-PBS, although the differences cannot be detected at conventional levels of statistical significance.²⁵ These results are consistent with the incentive structure in the scholarships, rather than primarily the additional income inducing the changes in behavior.

However, there was a critical difference in how the two types of scholarships were awarded that might also explain the larger impacts for the incentive scholarships: the incentive-based scholarships were paid directly to the students whereas the non-PBS was paid to the institutions. As such, the non-PBS may not have been as salient to the student; there is also the

²⁵ The results are similar if we compare the one-term, non-PBS scholarship impacts to the one-term, \$1,000 PBS and are available from the authors on request.

possibility that institutions at least partly off-set the non-PBS with reductions in other forms of financial aid. Further, one might interpret the fact that the PBS (column (2) of Table 6, Panel A) had a larger impact on enrollment than the non-PBS (columns (3) of Table 6, Panel A) as providing support for these alternative interpretations since one would expect the non-PBS to have a larger impact on ever enrolling in an institution than the PBS since the non-PBS was a guaranteed payment. The fact that the non-PBS does not appear to have affected enrollment may suggest the students were simply unaware of the award or that institutions off-set the award with reductions in other financial aid.

While we cannot completely rule out these alternative explanations, we suspect they do not explain the results for two reasons. First, the point estimate in column (3) suggests that the non-PBS had no impact on enrollment. Given the literature on the effect of college subsidies on enrollment this would only occur if the institutions completely off-set the non-PBS with reductions in other financial aid, which is unlikely given that institutions most often treat outside scholarship aid favorably, as detailed in footnote 8. Second, Deming and Dynarski (2010) conclude that the best estimates of the impact of educational subsidies on enrollment suggest that eligibility for \$1,000 in subsidy increases enrollment by about four percentage points, and the coefficient estimate for the impact of the non-PBS is not statistically different from this impact. In contrast, the estimated impact of the PBS scholarships on enrollment, when scaled by how much students actually received, is larger.

Taken together, we believe the evidence suggests an interpretation that incentives played a key role in changing student behavior. We note that this finding is also consistent with Scott-Clayton (2011) who studies the West Virginia PROMISE scholarship which is a merit scholarship with continuing eligibility that depends on meeting minimum credit and GPA

benchmarks and that had relatively large impacts on student educational attainment. As credit completion tended to be concentrated around the renewal thresholds, she concludes that the scholarship incentive was a key component for the success of the program.

D. Impacts by Type of Participant

Finally, we consider whether the impact of the incentive scholarships differs by type of participant. In particular, we hypothesize that the scholarships should have a larger impact for participants who have a lower marginal cost of time and those who are relatively more myopic. Because we do not directly observe these individual characteristics, we infer them based on background characteristics. Further, we rely on data from NYC because CA participants were not asked detailed background characteristics at baseline. The results are presented in Table 7.

In Panel A of Table 7, we estimate whether the incentive scholarships had a greater impact on those participants who did not have young children under the age of six on the assumption that parents of young children have less flexibility with their time given their parenting responsibilities which, in turn, raises the marginal cost of their study time. The coefficient estimates in column (1) represent the main effect of PBS eligibility; those in column (2) represent the interaction effect; the p-value on the interaction term is presented in column (3).

The estimated coefficient on the interaction term for all academic activities is positive in column (2) of Panel A in Table 7 indicating that the impact of the PBS was larger for those without children, as expected. Specifically, those without young children increased their time on academic activities by substantially more than those with young children. Notably, however, eligibility for a PBS generated a larger impact for those with young children compared to those without on the quality of educational input. This may not be surprising as one might expect that

those who find it costly to increase the quantity of their effort will try, instead, to increase the quality of that effort in order to reach the scholarship benchmark(s). In all cases the interaction term is not statistically significant at conventional levels due to large standard errors, but the pattern of coefficients is suggestive.

As a second exercise, we examine if the scholarships had a differential impact on a subgroup of students with whom policymakers are quite concerned – those who had completed 11 or fewer years of schooling before completing a GED or enrolling in (community) college. On the one hand, these individuals are arguably less well prepared for college because they did not complete their high school education. On the other hand, they are a reasonably large group, and they are getting a “second” chance at schooling. These results are reported in Panel B of Table 7.

Indeed, we find that incentives matter more for these second-chance students. For example, the program impact on time spent on educational activities was significantly larger for participants who dropped out of high school before 12th grade than for those who had completed more schooling; program impacts for non-academic activities and the quality of educational inputs were also larger. These differences are statistically significant at traditional levels. One possible explanation for this differential response is that those individuals who had dropped out of high school before enrolling in a community college are “myopic” in their time preferences as hypothesized by researchers such as Oreopoulos (2007). That is, the incentive may matter more for those who discount the future the greatest.²⁶

²⁶ We have also estimated differential impacts by an indicator of ex ante likelihood that a student will meet the benchmark (high school GPA or predicted probability using baseline characteristics) under the hypothesis that students predicted to be somewhat below the benchmark will have a larger response to the incentive than students who are highly likely to meet the benchmark. We find suggestive evidence that this is the case; however, the choice of the cut-off is not well-defined and the results are sensitive to this choice.

While we do not have direct measures of the marginal cost of time for participants or the rates at which they discount the future, we find evidence that the performance-based scholarships had larger impacts for some subgroups that could be explained by the incentive mechanisms largely working through hypothesized channels.

V. Conclusion

Although education policymakers have become increasingly interested in using incentives to improve educational outcomes, the evidence continues to generate, at best, small impacts, leading to the question of whether such incentives can actually change student effort toward their educational attainment as suggested by Becker's model of individual decision-making. As a whole, we find evidence consistent with this model: students eligible for performance-based scholarships increased effort in terms of the amount and quality of time spent on educational activities and decreased time spent on other activities. Further, it appears that such changes in behavior do not persist beyond eligibility for the performance-based scholarship suggesting that such incentives do not permanently change their cost of effort or their ability to transform effort into educational outcomes. And, students expected to be most responsive to the incentive – such as those with fewer time constraints and those who may be more myopic in their time preferences – likely were.

An important question arising from this study is why the larger incentive payments did not generate larger increases in effort. We offer a few potential explanations worthy of further consideration. First, the result may suggest that students need just a small prompt to encourage them to put more effort into their studies but that larger incentives are unnecessary. Further, it is possible that as the value of the incentive payment (external motivation) increases, students'

internal motivation declines at a faster rate such that negative impacts on intrinsic motivation increasingly moderate any positive impacts of the incentive on educational effort.²⁷ It may also be that students face constraints in their ability to change their effort level such that they are unable to change their behavior further, as suggested by the smaller impact on time on academic activities for participants who were also parents. Finally, these results could also be consistent with students not fully understanding their own “education production function,” i.e. how their own effort and ability will be transformed into academic outcomes like grades. While the students seem to understand that increases in effort are necessary to improve outcomes, they may overestimate their likelihood of meeting the benchmark and underestimate the marginal impact of effort on the probability of meeting the benchmark leading to suboptimal levels of effort.²⁸ While our data do not allow us to thoroughly understand why larger incentive payments did not generate larger changes in behavior, understanding why they did not would be important for the optimal design of incentive schemes to improve educational attainment.

More generally, this study highlights the potential benefits of better understanding student behavior in response to interventions to provide insights for future policy development. For example, if further study confirms that, indeed, those most likely to be constrained by time (such as parents) are less able to change the amount of time devoted to studies, then effective strategies to improve educational attainment among nontraditional students must recognize this reality among these students. Specifically, any efforts to improve their educational outcomes must also address constraints on their time, allowing them the ability to respond to the intervention or program, such as strategic scheduling and bundling of classes or more condensed

²⁷ The evidence available in our study on this last point does not suggest this is a promising explanation, but such evidence is far from conclusive.

²⁸ While this interpretation is appealing and we believe worthy of further consideration, it does suggest that we would expect larger scholarships to induce larger responses in the second semester after students have learned more about their own abilities and effectiveness at transforming effort into grades, which we do not find.

curricula. And, if this intervention was more effective for high school drop outs because of a difference in time preference, strategies for helping this population of students would be more effective if combined with systems that made rewards to increased academic effort more immediate.

Finally, while the evidence from this study of performance-based scholarships suggests modest impacts, such grants may nonetheless be a useful tool in postsecondary education policy as they appear to induce positive behavioral changes, and evidence from other similar studies, such as Barrow et al. (2014), suggests that even with small impacts on educational attainment, such relatively low-cost interventions may nonetheless be cost effective.

References

- Angrist, Joshua and Victor Lavy (2009). "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *The American Economic Review* 99:4, 1384-1414.
- 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, Philip Oreopoulos, and Tyler Williams (2012). "When Opportunity Knocks, Who Answers?: New Evidence on College Achievement Awards," MIT Working Paper, February.
- Baldi, Stephane, Ying Jin, Melanie Skemer, Patricia Green, and Deborah Herget Holly Xie (2007). "Highlight from PISA 2006: Performance of U.S. 15-Year-Old Students in Science and Mathematics Literacy in an International Context." National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Barrow, Lisa, Lashawn Richburg-Hayes, Cecilia Elena Rouse, and Thomas Brock (2014). "Paying for Performance: The Education Impacts of a Community College Scholarship Program for Low-income Adults." *Journal of Labor Economics*, forthcoming July 2014.
- Barrow, Lisa and Cecilia Elena Rouse (2006). U.S. Elementary and Secondary Schools: Equalizing Opportunity or Replicating Status Quo? *Future of Children Journal*, Vol. 16, No. 2, 2006, pp. 99-123.
- Barrow, Lisa and Cecilia Elena Rouse (2013). "Technical Report on the PBS Time Use Survey."
- Becker, Gary S., Human Capital and the Personal Distribution of Income: Ann Arbor, University of Michigan Press, 1967.
- Benabou, Roland and Jean Tirole (2003). "Intrinsic and Extrinsic Motivation," *The Review of Economic Studies* 70(3):489-520.
- Cha, Paulette and Reshma Patel (2010). "Rewarding Progress, Reducing Debt: Early Results from the Performance-Based Scholarship Demonstration in Ohio." New York: MDRC.
- Das, Jishnu, Quy-Toan Do and Berk Ozler (2005). "Reassessing Conditional Cash Transfer Programs," *The World Bank Research Observer* 20(1): 57-80.
- Deci, Edward L. (1975). *Intrinsic Motivation*. New York: Plenum.
- 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* 126, no. 6 (November): 627-668.

- Deci, Edward L. and Richard M. Ryan (1985). *Intrinsic Motivation and Self-determination in Human Behavior*. New York: Plenum.
- Deming, David and Susan Dynarski (2010). "Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor," in *Targeting Investments in Children: Fighting Poverty When Resources are Limited*, edited by Phil Levine and David Zimmerman, 208-302. Chicago: University of Chicago Press.
- Fryer, Roland G., Jr. (2011). "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *Quarterly Journal of Economics*. 126(4): 1755-1798.
- Gneezy, Uri and Aldo Rustichini (2000). "Pay Enough or Don't Pay at All," *The Quarterly Journal of Economics*. 115(3): 791-810.
- Huffman, David and Michael Bognanno (2012). "Performance Pay and Worker Intrinsic Motivation: Evidence from a Natural Field Experiment," Swarthmore College Mimeo.
- Jackson, C. Kirabo (2010a). "A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program" *Journal of Human Resources*. 45:3: 591-639.
- Jackson, C. Kirabo (2010b). "A Stitch in Time: The Effects of a Novel Incentive-Based High-School Intervention on College Outcomes," NBER Working Paper No. 15722.
- James, Jeannine M. and Richard Bolstein (1992). "Large Monetary Incentives and Their Effect on Mail Survey Response Rates," *Public Opinion Quarterly*. 56 no. 4: 442-453.
- Lau, Yan. "Tournament Structure and Effort in an Experimental Setting," Princeton University mimeo, June 2013.
- Midgley, Carol, Martin L. Maehr, Ludmila Z. Hruda, Eric Anderman, et al. 2000. "Manual for the Patterns of Adaptive Learning Scales." University of Michigan monograph.
- Miller, Cynthia, Melissa Binder, Vanessa Harris, and Kate Krause. 2011. "Staying on Track: Early Findings from a Performance-Based Scholarship Program at the University of New Mexico." New York: MDRC.
- National Commission on Excellence in Education. (1983). *A Nation at Risk: the Imperative for Educational Reform*. Washington: U.S. Dept. of Education.
- OECD (2011), "How Many Students Finish Tertiary Education?," in OECD, *Education at a Glance 2011: OECD Indicators*, OECD Publishing.
- Oreopoulos, Philip. (2007). "Do Dropouts Drop Out Too Soon? Wealth, Health, and Happiness from Compulsory Schooling." *Journal of Public Economics* 91: 2213-2229.

- Patel, Reshma, and Timothy Rudd. 2012. "Can Scholarships Alone Help Students Succeed? Lessons from Two New York City Community Colleges." New York: MDRC.
- Pintrich, Paul R. and Elisabeth V. De Groot. 1990. "Motivational and Self-Regulated Learning Components of Classroom Academic Performance." *Journal of Educational Psychology*, 82(1): 33-40.
- Pintrich, Paul R., David A. F. Smith, Teresa Garcia, and Wilbert J. McKeachie. 1991. *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*, Office of Educational Research and Improvement/Department of Education (OERI/ED) Technical Report No. 91-B-004.
- Plant, E. Ashby, K. Anders Ericsson, Len Hill, and Kia Asberg (2005). "Why Study Time Does Not Predict Grade Point Average Across College Students: Implications of Deliberate Practice for Academic Performance." *Contemporary Educational Psychology*, 30, pp. 96-116.
- Rawlings, Laura and Gloria Rubio (2005). "Evaluating the Impact of Conditional Cash Transfer Programs," *The World Bank Research Observer* 20(1): 29-55.
- Richburg-Hayes, Lashawn and Reshma Patel (forthcoming). *Interim Findings from the Performance-Based Scholarship Demonstration in California* (working title). New York: MDRC.
- Richburg-Hayes, Lashawn, Colleen Sommo, and Rashida Welbeck (2011). *Promoting Full-time Attendance Among Adults in Community College*. New York: MDRC.
- Scott-Clayton, Judith. (2011). "On Money and Motivation: A Quasi-Experimental Analysis A Quasi-Experimental Analysis of Financial Incentives for Colleges Achievement." *Journal of Human Resources*, vol. 46 (3): 614-646.
- Stinebrickner, Todd R. and Ralph Stinebrickner. (2007) "The Causal Effect of Studying on Academic Performance," NBER Working Paper #13341.
- Ware, Michelle and Reshma Patel. (2012) *Does More Money Matter?: An Introduction to the Performance-Based Scholarship Demonstration in California*. New York: MDRC.

Table 2: Total (Baseline) Sample Size by Site

Cohort	NYC		CA					
	PBS		Non-PBS (\$1,000)	PBS				
	\$1,300/term			\$ 500/term			\$1000/term	
	2 or 3 terms			1 term	2 terms	4 terms	1 term	2 terms
Fall 2008	368							
Spring 2009	514							
Fall 2009	619		483	484	447	468	468	460
Fall 2010			653	637	679	611	633	637
Total	1,501		1,136	1,121	1,126	1,079	1,101	1,097

Notes: The NYC sample size includes both those assigned to a 'regular' PBS and those assigned to a 'regular' PBS plus a summer PBS. Sample sizes for CA include 1,500 non-study individuals added to the control group.

Table 3: Characteristics of PBS Participants and First-year Students in the National Postsecondary Student Aid Study (NPSAS) of 2008

Characteristics	NYC	NPSAS 2-	CA	NPSAS All Types of
	PBS	Year Public Colleges	PBS	Institutions
	(1)	(2)	(3)	(4)
Age (years)	26.5	26.988	17.6	18.438
Age 17-18 (%)			96.7	60.1
Age 19-20 (%)			3.2	39.9
Age 21-35 (%)	99.9	96.2	0	
Female (%)	69.1	52.1	59.9	53.5
Race/ethnicity (%)				
Hispanic	44.3	21.6	63.2	15.4
Black	37.2	18.8	3.9	12.3
Asian	9.7	7.1	10.8	5.5
Native American	0.2	1.9	0.7	0.9
Other	1.0	1.8	0.3	4
Children				
Has any children (%)	47.8	46.9		2.4
Number of children	0.8	1.823		1.32
Household size		2.562		3.933
Financially dependent on parents (%)	1.3	11.9		94.7
Education				
Years since high school		6.958		0.135
Enrolled to complete certificate program	2.9	12.1		
Enrolled to transfer to 4-year college	43.1	32.1		
Highest degree completed				
GED	33.1	19.9		2.4
High school diploma	65.0	54.1		90.8
Technical certificate or AA	15.2	19.3		3.6
First family member to attend college (%)	32.9	47.0	54.8	28.7
Highest degree by either parent (%)				
Did not complete high school	24.2	13.4	36.4	4.1
High school diploma or equivalent	32.9	33.6	30.3	24.6
Some college including tech certificate	16.1	18.9		
Associate's or similar degree	6.4	8.5		
Some college including technical certificate, AA degree			22.3	27.4
4-year bachelor's degree or higher	20.3	25.6	11.1	43.9
U.S. citizen		88.6		95.5
Non-English spoken at home	54.6	19.6	63.0	12.0
Number of observations	1,501	250,997	6,660	2,660,060

Notes: Based on authors' calculations from MDRC data and data from the U.S. Department of Education's 2008 National Postsecondary Student Aid Study (NPSAS) of 2008. We limit the NPSAS data to first-time students, enrolled at any point from July 1 through December 31, 2007. For comparability with the PBS New York sample, in column (2) we limit the sample to students aged 22 to 36 attending public two-year colleges. For comparability with the PBS California sample, we include students aged 16-20 who are attending any type of institution in column (4). The NPSAS means and number of observations are weighted by the 2008 study weight.

Table 4a: Estimates of PBS Impact on Academic Outcomes

Variable	NYC			CA				
	Control Mean	PBS	Obs	Control Mean	PBS	Non-PBS	PBS = Non-PBS	Obs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ever Enrolled Postsecondary	0.922	-0.012 (0.023)	613	0.831	0.052*** (0.015)	0.004 (0.030)	0.124	2872
Hours on All Academics in Last 24 Hours	4.504	0.470 (0.314)	613	4.757	0.277 (0.174)	0.034 (0.345)	0.504	2874
Hours Studied in Past 7 Days	2.843	0.217 (0.204)	611	2.936	0.139 (0.098)	0.049 (0.195)	0.659	2871
Prepared for Last Class	0.810	0.026 (0.032)	606	0.736	0.073*** (0.018)	0.027 (0.036)	0.211	2861
Attended Most/All Classes in Past 7 Days	0.778	0.062* (0.032)	613	0.776	0.067*** (0.017)	0.028 (0.033)	0.267	2872
Academic Self-Efficacy	0.000	0.189** (0.078)	610	0.000	0.121*** (0.041)	0.024 (0.082)	0.260	2866
MSLQ Index	0.000	0.225*** (0.078)	613	0.000	0.224*** (0.042)	0.045 (0.083)	0.041	2871

Notes: Estimates obtained via OLS regressions including location-cohort fixed effects. Time-use variables refer to hours in past 24 hours unless otherwise noted. Column 7 shows the p-value for an F-test of the equality of the PBS and Non-PBS impacts. * indicates statistical significance at the 10% level; ** indicates statistical significance at the 5% level; and *** indicates statistical significance at the 1% level. MSLQ Index and Academic Self-Efficacy have been standardized to the relevant control group.

Table 4b: Estimates of PBS Impact on Quality of Non-academic Outcomes and Potential Unintended Consequences

Variable	NYC			CA				
	Control Mean	PBS	Obs	Control Mean	PBS	Non-PBS	PBS = Non-PBS	Obs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hours Worked in Last 24 Hours	2.496	0.096 (0.299)	613	0.750	0.026 (0.089)	0.101 (0.177)	0.687	2874
Hours Worked in Past 7 Days	14.953	0.671 (1.414)	605	4.928	-0.216 (0.399)	-0.142 (0.790)	0.928	2818
Hours on Household Production in Last 24 Hours	11.887	0.118 (0.352)	613	11.721	0.168 (0.147)	0.168 (0.291)	0.998	2874
Hours on Leisure in Last 24 Hours	5.080	-0.689** (0.302)	613	6.765	-0.482*** (0.160)	-0.302 (0.318)	0.591	2874
Times Out in Past 7 Days	0.761	-0.014 (0.084)	613	2.077	-0.124** (0.059)	-0.165 (0.118)	0.746	2863
Strongly Agree/Agree to Take Challenging Classes	0.451	0.024 (0.041)	607	0.385	0.058*** (0.021)	0.037 (0.041)	0.638	2856
Ever Felt Had to Cheat	0.176	-0.027 (0.030)	611	0.349	-0.106*** (0.019)	-0.106*** (0.039)	0.989	2854
External Motivation	0.000	0.031 (0.088)	558	0.000	0.077* (0.044)	-0.132 (0.090)	0.025	2417
Internal Motivation	0.000	0.195*** (0.076)	560	0.000	0.019 (0.045)	-0.124 (0.092)	0.136	2419
Ever Asked for Regrade	0.262	-0.018 (0.036)	609	0.197	0.006 (0.017)	-0.087*** (0.033)	0.008	2860
Very Satisfied/Satisfied with Life	0.494	0.070* (0.041)	608	0.624	0.010 (0.020)	-0.002 (0.041)	0.781	2850

Notes: Estimates obtained via OLS regressions including location-cohort fixed effects. Column 7 shows the p-value for an F-test of the equality of the PBS and Non-PBS impacts. * indicates statistical significance at the 10% level; ** indicates statistical significance at the 5% level; and *** indicates statistical significance at the 1% level. The indices of Internal Motivation and External Motivation have been standardized to the relevant control group.

Table 5: Index Estimates of PBS-Impact for NYC & CA

	NYC		CA			
	PBS Impact	Obs	PBS Impact	Non-PBS Impact	P-Val for PBS=Non-PBS	Obs
	(1)	(2)	(3)	(4)	(5)	(6)
All Academic Activities	0.106** (0.051)	613	0.113*** (0.027)	0.039 (0.056)	0.203	2874
Quality of Educational Input	0.207*** (0.068)	613	0.173*** (0.035)	0.034 (0.075)	0.077*	2872
Non-Academic Activities	-0.021 (0.030)	613	-0.035* (0.018)	-0.023 (0.034)	0.730	2874
Unintended Consequences	-0.077** (0.033)	613	-0.048*** (0.018)	-0.087** (0.037)	0.315	2874

Notes: Estimates are indexed estimates obtained via the seemingly unrelated regression (SUR) strategy discussed in the paper. All regressions include location-cohort fixed-effects. In the "Unintended Consequences" Index, Internal Motivation, Agree to Take Challenging Classes & Satisfied with Life are adjusted so that a negative indicates a "good" outcome. * indicates statistical significance at the 10% level; ** indicates statistical significance at the 5% level; and *** indicates statistical significance at the 1% level.

Table 6: PBS Impact in CA by Scholarship Size and Length

Panel A: Impact by Scholarship Size and Incentives in the First Semester						
	\$500/T	\$1000/T	Non-PBS	\$500/T =\$1000/T	Non-PBS =\$1000/T	Obs
	(1)	(2)	(3)	(4)	(5)	(6)
Ever Enrolled	0.072*** (0.021)	0.039** (0.018)	0.004 (0.030)	0.188	0.295	2872
Currently Enrolled	0.071*** (0.023)	0.049** (0.020)	0.028 (0.032)	0.430	0.554	2873
All Academic Activities	0.120*** (0.036)	0.108*** (0.033)	0.039 (0.056)	0.772	0.260	2874
Quality of Educational Input	0.191*** (0.048)	0.160*** (0.043)	0.034 (0.075)	0.597	0.125	2872
Non-Academic Activities	-0.026 (0.025)	-0.041* (0.022)	-0.023 (0.034)	0.630	0.632	2874
Unintended Consequences	-0.055** (0.026)	-0.044* (0.022)	-0.087** (0.037)	0.731	0.288	2874
Panel B: Impact by Scholarship Length in the Second Semester						
	1 Term	2+ Terms	Non-PBS	1 Term = 2 Terms	Non-PBS = 1 Term	Obs
	(1)	(2)	(3)	(4)	(5)	(6)
Ever Enrolled	0.032 (0.028)	0.013 (0.015)	0.003 (0.029)	0.534	0.470	2742
Currently Enrolled	0.003 (0.032)	0.016 (0.017)	0.015 (0.033)	0.704	0.795	2740
All Academic Activities	0.038 (0.057)	0.082*** (0.030)	-0.017 (0.057)	0.468	0.477	2743
Quality of Educational Input	0.139* (0.078)	0.125*** (0.038)	0.108 (0.076)	0.861	0.766	2742
Non-Academic Activities	0.029 (0.039)	-0.066*** (0.019)	-0.007 (0.034)	0.021	0.470	2743
Unintended Consequences	-0.063* (0.037)	-0.067*** (0.019)	-0.104*** (0.036)	0.923	0.420	2742

Notes: Estimates for enrollment obtained via OLS regressions. Other estimates obtained via the SUR strategy discussed in the paper. Regressions include cohort-location fixed effects. Column 5 shows the p-value for an F-test of the equality of the PBS and Non-PBS impacts. In the "Unintended Consequences" Index, Internal Motivation, Agree to Take Challenging Classes, and Satisfied with Life are adjusted so that a negative indicates a "good" outcome. * indicates statistical significance at the 10% level; ** indicates statistical significance at the 5% level; and *** indicates statistical significance at the 1% level.

Table 7: PBS Impact in NYC by Respondent Characteristics

	Panel A: PBS Impact by Parental Status		
	PBS (1)	PBS x No Young Child (2)	P-Val of Interaction (3)
All Academic Activities	0.038 (0.080)	0.086 (0.105)	0.410
Quality of Educational Input	0.250** (0.117)	-0.091 (0.144)	0.528
Non-Academic Activities	-0.046 (0.049)	0.047 (0.062)	0.451
Unintended Consequences	-0.121** (0.052)	0.070 (0.068)	0.301
	Panel B: PBS Impact by Previous Education Attainment		
	PBS (1)	PBS x ≤ 11 yrs Education (2)	P-Val of Interaction (3)
All Academic Activities	0.054 (0.064)	0.200* (0.109)	0.065
Quality of Educational Input	0.093 (0.084)	0.349** (0.152)	0.022
Non-Academic Activities	0.007 (0.037)	-0.107* (0.065)	0.101
Unintended Consequences	-0.033 (0.042)	-0.091 (0.071)	0.205

Notes: Estimates are indexed estimates obtained via the SUR strategy discussed in the paper. Regressions also include an indicator for parental status/low educational attainment and cohort-location fixed effects. In the "Unintended Consequences" Index, Internal Motivation, Agree to Take Challenging Classes, and Satisfied with Life are adjusted so that a negative indicates a "good" outcome. * indicates statistical significance at the 10% level; ** indicates statistical significance at the 5% level; and *** indicates statistical significance at the 1% level. There are 609 observations in Panel A and 570 observations in Panel B.

Appendix Table A1a: Randomization of Program and Control Groups at NYC Sites

Baseline Characteristic (%)	Random Assignment		p-value of difference	N
	Program Group	Control Group		
Age (years)	26.5	26.6	0.713	1501
Marital Status				
Married, living with spouse	11.1	13.8	0.129	1384
Married, not living with spouse	7.4	7.2	0.916	1384
Not married, living with partner	12.1	10.1	0.206	1384
Single	69.3	68.9	0.87	1384
Female	69.8	68.4	0.568	1501
No children under six	69.1	65.5	0.119	1492
Race/ethnicity				
Hispanic	44.4	44.2	0.995	1468
Black	36.2	38.2	0.419	1468
White	6.3	5.9	0.736	1468
Asian	10.3	9.0	0.381	1468
Native American	0.1	0.3	0.569	1468
Other	1.0	1.1	0.791	1468
Multi-racial	1.8	1.2	0.4	1468
Race not reported	2.5	1.9	0.395	1501
Household receiving benefits				
Receiving any government benefit	42.2	43.9	0.528	1321
Receiving unemployment insurance	7.7	11.5	0.017	1321
Household receiving SSI	6.6	6.1	0.703	1321
Household receiving TANF	9.2	6.9	0.123	1321
Household receiving food stamps	30.1	30.2	0.959	1321
Public housing or section 8 housing	10.6	10.6	0.999	1321
Financially dependent on parents	1.0	1.7	0.231	1435
Currently employed	56.5	55.4	0.648	1446
Years since HS (years)	6.8	6.9	0.851	1375
High school diploma or GED	96.5	96.5	0.884	1470
Technical certificate	11.8	14.8	0.082	1470
Last attended 11th grade or lower	29.5	31.7	0.345	1408
First family member to attend college	34.5	31.2	0.194	1454
Main reason for enrolling in college				
Complete certificate program	3.0	2.8	0.89	1478
Obtain Associate's degree	48.5	52.7	0.104	1478
Transfer to four-year college	46.0	40.0	0.017	1478
Obtain job skills	2.8	3.7	0.375	1478
Other reason	1.2	2.0	0.204	1478
Primary language				
English	45.5	45.2	0.885	1487

Spanish	29.0	29.8	0.717	1487
Other language	25.3	24.8	0.836	1487

Notes: Calculations using Baseline Information Form (BIF) data. Means have been adjusted by research cohort and site. An omnibus F-test of whether baseline characteristics jointly predict research group status yielded a p-value of 0.68. Distributions may not add to 100 percent because of rounding. Respondents that reported being Hispanic/Latino and also reported a race are included only in the Hispanic category. Respondents that are not coded as Hispanic and chose more than one race are coded as multi-racial.

Appendix Table A1b: Randomization of Program and Control Groups at CA Sites

Baseline Characteristic (%)	Random Assignment		p-value of difference	N
	Program Group	Control Group		
Age (years)	17.6	17.6	0.224	6660
Female	60.4	59.7	0.611	6659
Race/ethnicity				
Hispanic	63.2	63.2	0.981	6597
Black	3.4	4.1	0.208	6597
White	18.2	18.7	0.538	6597
Asian	10.5	10.8	0.601	6597
Native American	0.7	0.7	0.996	6597
Other	0.4	0.3	0.593	6597
Multi-racial	3.6	2.2	0.001	6597
Race not reported	0.7	0.4	0.469	2810
Highest degree by either parent				
No high school diploma	36.7	36.2	0.639	6541
High school diploma/GED	29.7	30.5	0.536	6541
Associate's or similar degree	22.7	22.1	0.532	6541
Bachelor's degree	10.6	11.1	0.56	6541
First family member to attend college	56.5	54.7	0.185	6612
Primary language				
English	37.9	36.5	0.248	6617
Spanish	50.7	51.5	0.458	6617
Other language	11.5	11.8	0.601	6617

Notes: Calculations using Baseline Information Form (BIF) data. The means have been adjusted by research cohort and workshop region. An omnibus F-test of whether baseline characteristics jointly predict research group status yielded a p-value of 0.476. Distributions may not add to 100 percent because of rounding. Respondents who reported being Hispanic/Latino and also reported a race are included only in the Hispanic category. Respondents who are not coded as Hispanic and chose more than one race are coded as multi-racial.

Appendix Table A2a: Random Assignment of Program and Control Groups at NYC sites, Analysis Sample

Baseline Characteristic (%)	Random Assignment		p-value of difference	N
	Program Group	Control Group		
Age (years)	26.3	26.7	0.375	613
Marital Status				
Married, living with spouse	12.8	13.3	0.825	559
Married, not living with spouse	7.2	9.6	0.326	559
Not married, living with partner	12.8	7.9	0.068	559
Single	67.1	69.0	0.633	559
Female	71.9	75.5	0.326	613
No children under six	69.0	61.7	0.061	609
Race/ethnicity				
Hispanic	43.4	44.4	0.81	598
Black	38.4	40.5	0.582	598
White	5.6	5.0	0.741	598
Asian	10.0	6.1	0.087	598
Native American	0.0	0.0		598
Other	0.9	1.2	0.722	598
Multi-racial	1.7	2.8	0.405	598
Race not reported	2.6	2.3	0.814	613
Household receiving benefits				
Receiving any government benefit	47.2	51.2	0.36	543
Receiving unemployment insurance	9.1	16.3	0.009	543
Household receiving SSI	6.4	6.1	0.885	543
Household receiving TANF	8.8	8.9	0.961	543
Household receiving food stamps	33.2	36.5	0.43	543
Public housing or section 8 housing	12.6	12.0	0.852	543
Financially dependent on parents	0.9	1.6	0.409	592
Currently employed	53.2	52.0	0.764	589
Years since HS (years)	6.7	6.8	0.816	566
High school diploma or GED	96.5	97.3	0.62	600
Technical certificate	13.5	13.3	0.972	600
Last attended 11th grade or lower	29.7	32.7	0.435	570
First family member to attend college	36.2	29.2	0.075	597
Main reason for enrolling in college				
Complete certificate program	3.2	2.7	0.763	605
Obtain Associate's degree	50.2	53.7	0.381	605
Transfer to four-year college	45.2	39.4	0.15	605
Obtain job skills	2.8	3.6	0.598	605
Other reason	1.4	0.8	0.432	605
Primary language				

English	43.4	43.4	0.986	606
Spanish	31.3	32.5	0.768	606
Other language	25.2	24.1	0.745	606

Notes: Calculations using Baseline Information Form (BIF) data. The means have been adjusted by research cohort and site. An omnibus F-test of whether baseline characteristics jointly predict research group status yielded a p-value of 0.63. Distributions may not add to 100 percent because of rounding. Respondents who reported being Hispanic/Latino and also reported a race are included only in the Hispanic category. Respondents who are not coded as Hispanic and chose more than one race are coded as multi-racial.

Appendix Table A2b: Randomization of Program and Control Groups at CA Sites, Analysis Sample

Baseline Characteristic (%)	Random Assignment		p-value of difference	N
	Program Group	Control Group		
Age (years)	17.6	17.6	0.116	2874
Female	62.7	63.9	0.523	2874
Race/ethnicity				
Hispanic	62.0	62.0	0.982	2847
Black	3.4	3.3	0.875	2847
White	18.0	18.7	0.493	2847
Asian	12.1	12.8	0.547	2847
Native American	0.4	0.4	0.795	2847
Other	0.6	0.2	0.043	2847
Multi-racial	3.3	2.3	0.127	2847
Race not reported	0.8	1.0	0.628	2874
Highest degree by either parent				
No high school diploma	37.5	38.0	0.813	2837
High school diploma/GED	28.3	28.7	0.832	2837
Associate's or similar degree	23.5	21.5	0.246	2837
Bachelor's degree	10.6	11.6	0.374	2837
First family member to attend college	54.4	55.0	0.782	2860
Primary language				
English	35.0	34.5	0.762	2860
Spanish	51.4	51.9	0.746	2860
Other language	13.5	13.3	0.967	2860

Notes: Calculations using Baseline Information Form (BIF) data. The means have been adjusted by research cohort and workshop region. An omnibus F-test of whether baseline characteristics jointly predict research group status yielded a p-value of 0.72. Distributions may not add to 100 percent because of rounding. Respondents that reported being Hispanic/Latino and also reported a race are included only in the Hispanic category. Respondents that are not coded as Hispanic and chose more than one race are coded as multi-racial.

Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

Comment on “Letting Different Views about Business Cycles Compete” <i>Jonas D.M. Fisher</i>	WP-10-01
Macroeconomic Implications of Agglomeration <i>Morris A. Davis, Jonas D.M. Fisher and Toni M. Whited</i>	WP-10-02
Accounting for non-annuitization <i>Svetlana Pashchenko</i>	WP-10-03
Robustness and Macroeconomic Policy <i>Gadi Barlevy</i>	WP-10-04
Benefits of Relationship Banking: Evidence from Consumer Credit Markets <i>Sumit Agarwal, Souphala Chomsisengphet, Chunlin Liu, and Nicholas S. Souleles</i>	WP-10-05
The Effect of Sales Tax Holidays on Household Consumption Patterns <i>Nathan Marwell and Leslie McGranahan</i>	WP-10-06
Gathering Insights on the Forest from the Trees: A New Metric for Financial Conditions <i>Scott Brave and R. Andrew Butters</i>	WP-10-07
Identification of Models of the Labor Market <i>Eric French and Christopher Taber</i>	WP-10-08
Public Pensions and Labor Supply Over the Life Cycle <i>Eric French and John Jones</i>	WP-10-09
Explaining Asset Pricing Puzzles Associated with the 1987 Market Crash <i>Luca Benzoni, Pierre Collin-Dufresne, and Robert S. Goldstein</i>	WP-10-10
Prenatal Sex Selection and Girls’ Well-Being: Evidence from India <i>Luoqia Hu and Analía Schlosser</i>	WP-10-11
Mortgage Choices and Housing Speculation <i>Gadi Barlevy and Jonas D.M. Fisher</i>	WP-10-12
Did Adhering to the Gold Standard Reduce the Cost of Capital? <i>Ron Alquist and Benjamin Chabot</i>	WP-10-13
Introduction to the <i>Macroeconomic Dynamics</i> : Special issues on money, credit, and liquidity <i>Ed Nosal, Christopher Waller, and Randall Wright</i>	WP-10-14
Summer Workshop on Money, Banking, Payments and Finance: An Overview <i>Ed Nosal and Randall Wright</i>	WP-10-15
Cognitive Abilities and Household Financial Decision Making <i>Sumit Agarwal and Bhashkar Mazumder</i>	WP-10-16

Working Paper Series *(continued)*

Complex Mortgages <i>Gene Amromin, Jennifer Huang, Clemens Sialm, and Edward Zhong</i>	WP-10-17
The Role of Housing in Labor Reallocation <i>Morris Davis, Jonas Fisher, and Marcelo Veracierto</i>	WP-10-18
Why Do Banks Reward their Customers to Use their Credit Cards? <i>Sumit Agarwal, Sujit Chakravorti, and Anna Lunn</i>	WP-10-19
The impact of the originate-to-distribute model on banks before and during the financial crisis <i>Richard J. Rosen</i>	WP-10-20
Simple Markov-Perfect Industry Dynamics <i>Jaap H. Abbring, Jeffrey R. Campbell, and Nan Yang</i>	WP-10-21
Commodity Money with Frequent Search <i>Ezra Oberfield and Nicholas Trachter</i>	WP-10-22
Corporate Average Fuel Economy Standards and the Market for New Vehicles <i>Thomas Klier and Joshua Linn</i>	WP-11-01
The Role of Securitization in Mortgage Renegotiation <i>Sumit Agarwal, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff</i>	WP-11-02
Market-Based Loss Mitigation Practices for Troubled Mortgages Following the Financial Crisis <i>Sumit Agarwal, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff</i>	WP-11-03
Federal Reserve Policies and Financial Market Conditions During the Crisis <i>Scott A. Brave and Hesna Genay</i>	WP-11-04
The Financial Labor Supply Accelerator <i>Jeffrey R. Campbell and Zvi Hercowitz</i>	WP-11-05
Survival and long-run dynamics with heterogeneous beliefs under recursive preferences <i>Jaroslav Borovička</i>	WP-11-06
A Leverage-based Model of Speculative Bubbles (Revised) <i>Gadi Barlevy</i>	WP-11-07
Estimation of Panel Data Regression Models with Two-Sided Censoring or Truncation <i>Sule Alan, Bo E. Honoré, LuoJia Hu, and Søren Leth-Petersen</i>	WP-11-08
Fertility Transitions Along the Extensive and Intensive Margins <i>Daniel Aaronson, Fabian Lange, and Bhashkar Mazumder</i>	WP-11-09
Black-White Differences in Intergenerational Economic Mobility in the US <i>Bhashkar Mazumder</i>	WP-11-10

Working Paper Series *(continued)*

Can Standard Preferences Explain the Prices of Out-of-the-Money S&P 500 Put Options? <i>Luca Benzoni, Pierre Collin-Dufresne, and Robert S. Goldstein</i>	WP-11-11
Business Networks, Production Chains, and Productivity: A Theory of Input-Output Architecture <i>Ezra Oberfield</i>	WP-11-12
Equilibrium Bank Runs Revisited <i>Ed Nosal</i>	WP-11-13
Are Covered Bonds a Substitute for Mortgage-Backed Securities? <i>Santiago Carbó-Valverde, Richard J. Rosen, and Francisco Rodríguez-Fernández</i>	WP-11-14
The Cost of Banking Panics in an Age before “Too Big to Fail” <i>Benjamin Chabot</i>	WP-11-15
Import Protection, Business Cycles, and Exchange Rates: Evidence from the Great Recession <i>Chad P. Bown and Meredith A. Crowley</i>	WP-11-16
Examining Macroeconomic Models through the Lens of Asset Pricing <i>Jaroslav Borovička and Lars Peter Hansen</i>	WP-12-01
The Chicago Fed DSGE Model <i>Scott A. Brave, Jeffrey R. Campbell, Jonas D.M. Fisher, and Alejandro Justiniano</i>	WP-12-02
Macroeconomic Effects of Federal Reserve Forward Guidance <i>Jeffrey R. Campbell, Charles L. Evans, Jonas D.M. Fisher, and Alejandro Justiniano</i>	WP-12-03
Modeling Credit Contagion via the Updating of Fragile Beliefs <i>Luca Benzoni, Pierre Collin-Dufresne, Robert S. Goldstein, and Jean Helwege</i>	WP-12-04
Signaling Effects of Monetary Policy <i>Leonardo Melosi</i>	WP-12-05
Empirical Research on Sovereign Debt and Default <i>Michael Tomz and Mark L. J. Wright</i>	WP-12-06
Credit Risk and Disaster Risk <i>François Gourio</i>	WP-12-07
From the Horse’s Mouth: How do Investor Expectations of Risk and Return Vary with Economic Conditions? <i>Gene Amromin and Steven A. Sharpe</i>	WP-12-08
Using Vehicle Taxes To Reduce Carbon Dioxide Emissions Rates of New Passenger Vehicles: Evidence from France, Germany, and Sweden <i>Thomas Klier and Joshua Linn</i>	WP-12-09
Spending Responses to State Sales Tax Holidays <i>Sumit Agarwal and Leslie McGranahan</i>	WP-12-10

Working Paper Series *(continued)*

Micro Data and Macro Technology <i>Ezra Oberfield and Devesh Raval</i>	WP-12-11
The Effect of Disability Insurance Receipt on Labor Supply: A Dynamic Analysis <i>Eric French and Jae Song</i>	WP-12-12
Medicaid Insurance in Old Age <i>Mariacristina De Nardi, Eric French, and John Bailey Jones</i>	WP-12-13
Fetal Origins and Parental Responses <i>Douglas Almond and Bhashkar Mazumder</i>	WP-12-14
Repos, Fire Sales, and Bankruptcy Policy <i>Gaetano Antinolfi, Francesca Carapella, Charles Kahn, Antoine Martin, David Mills, and Ed Nosal</i>	WP-12-15
Speculative Runs on Interest Rate Pegs The Frictionless Case <i>Marco Bassetto and Christopher Phelan</i>	WP-12-16
Institutions, the Cost of Capital, and Long-Run Economic Growth: Evidence from the 19th Century Capital Market <i>Ron Alquist and Ben Chabot</i>	WP-12-17
Emerging Economies, Trade Policy, and Macroeconomic Shocks <i>Chad P. Bown and Meredith A. Crowley</i>	WP-12-18
The Urban Density Premium across Establishments <i>R. Jason Faberman and Matthew Freedman</i>	WP-13-01
Why Do Borrowers Make Mortgage Refinancing Mistakes? <i>Sumit Agarwal, Richard J. Rosen, and Vincent Yao</i>	WP-13-02
Bank Panics, Government Guarantees, and the Long-Run Size of the Financial Sector: Evidence from Free-Banking America <i>Benjamin Chabot and Charles C. Moul</i>	WP-13-03
Fiscal Consequences of Paying Interest on Reserves <i>Marco Bassetto and Todd Messer</i>	WP-13-04
Properties of the Vacancy Statistic in the Discrete Circle Covering Problem <i>Gadi Barlevy and H. N. Nagaraja</i>	WP-13-05
Credit Crunches and Credit Allocation in a Model of Entrepreneurship <i>Marco Bassetto, Marco Cagetti, and Mariacristina De Nardi</i>	WP-13-06

Working Paper Series *(continued)*

Financial Incentives and Educational Investment:
The Impact of Performance-Based Scholarships on Student Time Use
Lisa Barrow and Cecilia Elena Rouse

WP-13-07