

FINANCIAL INCENTIVES AND EDUCATIONAL INVESTMENT: THE IMPACT OF PERFORMANCE-BASED SCHOLARSHIPS ON STUDENT TIME USE

Lisa Barrow

(corresponding author)
Senior Economist and
Research Advisor
Federal Reserve Bank of
Chicago
Chicago, IL 60604
lbarrow@frbchi.org

Cecilia Elena Rouse

Katzman-Ernst Professor in
the Economics of Education
Woodrow Wilson School of
Public and International
Affairs
Princeton University
Princeton, NJ 08544
rouse@princeton.edu

Abstract

We evaluate the effect of performance-based scholarship programs for postsecondary students on student time use and effort. We find evidence that financial incentives induced students to devote more time and effort to educational activities and allocate less time to other activities. Incentives did not generate impacts after eligibility ended and did not decrease students' interest or enjoyment in learning. Evidence also suggests that students were motivated more by the incentives than simply the effect of additional money. A remaining puzzle is that larger scholarships did not generate larger responses in terms of effort.

https://doi.org/10.1162/edfp_a_00228

© 2018 Association for Education Finance and Policy

1. INTRODUCTION

Educators have long worried about relatively low levels of educational performance among U.S. students.¹ Although on international tests of mathematics and science fourth- and eighth-grade elementary school students in the United States perform comparably to their peers in other developed countries (Provasnik et al. 2012), fifteen-year-old students in the United States perform poorly compared with their international peers in these same subjects (Kelly et al. 2013). Further, at the postsecondary level there is increasing concern that whereas the United States has one of the highest rates of college attendance in the world, rates of college completion lag those of other countries (OECD 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 toward their studies. Fundamentally, if students are to succeed in college, they need to master the material taught in a number of courses in order to earn credits. This takes effort. Students put effort into their education in at least two dimensions—attending class and studying—and students can increase effort in these dimensions by attending class more frequently, increasing the time they spend studying, or improving the quality of their time spent studying or in class. Little is known about the causal impact of effort on educational outcomes, both because effort is not easily measured and because it is a choice variable. Observational studies suggest that class attendance has a positive impact on student grades (see, e.g., Romer 1993). However, the best evidence of a causal impact of effort on student grades comes from Stinebrickner and Stinebrickner (2008), who use an instrumental variables strategy and find that increased time spent studying has a positive and statistically significant impact on student grades. Specifically, they estimate that one additional hour spent studying per week increases first-semester grade point averages (GPAs) by 0.36 GPA points, the equivalent of a 1.1 standard deviation increase in ACT test scores. Therefore, increasing students' effort toward their studies leads to improved educational outcomes. Further, the large decline in time devoted to studies documented by Babcock and Marks (2011; from 40 hours per week for full-time students in 1961 to 27 hours per week for full-time students in 2004) suggests that there is scope for students to increase time spent investing in their education.

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, Do, and Ozler 2005 and Rawlings and Rubio 2005), U.S. educators have implemented programs in which students are paid for achieving benchmarks such as a minimum GPA or for reading a minimum number of books. The U.S. strategies are based on the belief that current payoffs 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 payoffs, these incentive-based strategies are designed to provide students with a bigger incentive to work hard.

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 (U.S. National Commission on Excellence in Education 1983).

Unfortunately, the evidence to-date of such efforts has yielded somewhat mixed impacts on student achievement. For example, Jackson (2010a, b) finds some evidence that the Advanced Placement Incentive Program in Texas—which 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 postsecondary school grades. These results are similar to those reported by Angrist and Lavy (2009) and Kremer, Miguel, and Thornton (2009) from high school incentive programs in Israel and Kenya, respectively, which reward students for exam scores. 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, but rewarding them for education outcomes (such as grades and test scores) does not. One might think, however, that the impacts of incentives for younger students may be different from the impacts of incentives for postsecondary students.

At the postsecondary level, estimated impacts of financial incentives have been modest. For example, Angrist, Lang, and Oreopoulos (2009) and Angrist, Oreopoulos, and Williams (2014) report small impacts on grades at a four-year college in Canada, although impacts may be larger for some subgroups.² Barrow et al. (2014) report positive impacts on enrollment and credits earned for a program aimed at low-income adults attending community college in the United States, 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 Patel et al. 2013). Results are somewhat mixed from studies of the larger state-run merit aid programs that target high school students meeting minimum qualification requirements and then require students to meet benchmarks for renewal. Dynarski (2008) finds a 3 percentage point increase in college completion rates from the Arkansas and Georgia merit aid programs; however, Sjoquist and Winters (2012) find no impact on the college completion rates for these same programs when using a larger sample.³ Furthermore, Cornwell, Lee, and Mustard (2005) find that the Georgia program increased time to degree. Scott-Clayton (2011) finds relatively sizable impacts of the West Virginia PROMISE program on the probability of earning at least thirty credits and maintaining the benchmark GPA in each of the first three years of college, as well as a 2 percentage point increase in the Bachelor of Arts (BA) completion rate relative to a base BA completion rate of 21.5 percent. Importantly, the PROMISE program had minimum credit requirements in addition to GPA requirements, and Scott-Clayton (2011) finds that students respond to the incentive more strongly during the first three years of PROMISE receipt, when meeting the benchmark is required for renewal.⁴

2. Specifically, Angrist, Lang, and Oreopoulos (2009) report larger impacts for women, although the subgroup result is not replicated in Angrist, Oreopoulos, and Williams (2014). These authors report larger effects among those aware of the program rules.
3. See Angrist, Oreopoulos, and Williams (2014) for a more extensive discussion of this literature.
4. Lindo, Sanders, and Oreopoulos (2010) and Casey et al. (2018) study the impact of academic probation on subsequent academic outcomes with mixed results, including some evidence of improvements due to strategic behavior. Schudde and Scott-Clayton (2016) similarly find mixed results when examining how failing to meet satisfactory academic performance standards affects subsequent academic outcomes for Pell Grant recipients.

The fact that impacts of financial incentives on academic outcomes at the postsecondary level have been mostly modest raises the question of why. For example, can educational investments be effectively influenced through the use of incentives in the sense that they actually change student behavior? Alternatively, are these small, positive results a statistical anomaly, do they reflect the provision of additional income rather than the incentive structure, or do they represent changes along other dimensions, such as taking easier classes? And finally, can the variation in estimated impacts be explained by the size of the scholarships?

In this paper, we evaluate the effect of a performance-based scholarship (PBS) program for postsecondary students on a variety of outcomes, but especially on student effort as reflected in time use. High school seniors in California 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 PBSs on student educational effort, we surveyed participants about time use over the prior week and implemented a time diary survey.

Our paper contributes to the literature on incentives in education along a number of dimensions. First, it provides the initial estimates of the impact of these types of scholarships on postsecondary school enrollment for high school seniors who have yet to enroll in college.⁵ Second, data on time use allow us to get inside the “black box” to understand how postsecondary students may have changed their time allocation in response to the incentives. Further, variation in the incentive structure of the scholarships allows us to test a variety of hypotheses about the impacts of incentive payments among college-aged students in the United States. Specifically, we test whether larger payments or longer payments differentially affect outcomes. Finally, we provide some evidence on whether students respond to the incentives or the increased income.

We find that high school students eligible for a PBS were more likely to enroll in college the following semester. Further, those eligible for a scholarship 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 the incentives did not affect student behavior—either positively or negatively—after incentive payments were removed, suggesting that students were motivated by the incentives provided by the scholarships rather than simply the additional money. We note that the surveys upon which these results are based only yielded response rates between 43 percent (second semester) and 45 percent (first semester), raising the concern that the results are driven by selection bias. We have implemented a variety of tests to assess this possibility and ultimately conclude the estimates are internally valid. We base this assessment on the fact that omnibus F-tests of the difference in means of background characteristics at baseline by treatment/control status for our analysis sample are not statistically significant at conventional levels. Further, our estimates are largely unchanged if we include baseline characteristics as controls or use an inverse-probability weighting estimator to correct for selection.

5. MDRC has since released their interim findings on performance-based scholarships in California (see Richburg-Hayes et al. 2015). Their reported impacts on enrollment in California are largely consistent with those reported here.

Unfortunately, Lee bounds estimates (Lee 2009) are quite wide and largely uninformative. However, when we assess selection on unobservables using the strategy suggested by Altonji, Elder, and Taber (2005), we conclude that this is unlikely to overturn our primary findings. Keeping this in mind, overall our findings indicate that well-designed incentives can induce postsecondary 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 2). We describe the intervention studied, the data, and sample characteristics of program participants in section 3. The estimation strategy and results are presented in section 4, and section 5 concludes.

2. 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. Investing in education is costly in terms of effort. Benefits in our setting include the present discounted value of the earnings increase associated with additional college credits (net of tuition and other costs) as well as incentive payments for eligible students meeting the minimum GPA benchmark. Students maximize utility by maximizing the net expected benefit of effort. If the marginal benefit is relatively low or the marginal costs are relatively high a student may not enroll or continue in college. For a student who enrolls, an increase in the payoff to meeting the benchmark will lead to an increase in effort toward her studies. Likewise, a fall in the payoff will lead to a decrease in effort.⁶

Although this is a simple static model of optimal choice of effort, there are potential dynamic effects that we will consider and test in addition to 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 regardless of whether the student passes her classes once she gets there. In the theoretical framework outlined 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 version of the model, future scholarship receipt may depend on meeting the 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 toward schoolwork in the current semester by increasing the current payoff. 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, the direct effect of increasing the size of the

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

incentive payment in the current semester is a contemporaneous increase in effort but no impact on effort in future semesters.⁷ 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.⁸ Similarly, if increased effort today teaches students how to study more effectively, then PBS eligibility could also have a lasting positive impact on student outcomes.

At the same time, cognitive psychologists worry that although incentive payments may motivate students to do better in the short term, students may be motivated for the “wrong” 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 PB-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 although such scholarships may increase external motivation, they may decrease internal motivation (e.g., Deci, Koestner, and Ryan 1999). 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, forthcoming; Benabou and Tirole 2003). 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).

Finally, although the intention of a PBS 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 they would meet the minimum GPA benchmark. Other unintended consequences could include cheating on exams, asking professors to regrade tests and/or papers, or taking easier classes.

3. 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 eight institutions and at a state-wide

7. Ariely et al. (2009) note that although it is generally accepted that temporary increases in incentive payments lead to greater effort, there are several studies in which larger incentive payments (or higher effective wages) do not lead to increased effort.

8. 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.

organization in 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” survey module we implemented in California (CA).⁹ Our Time Use survey was coordinated with MDRC but was outside the scope of their project. In this demonstration, the scholarships supplemented any other financial aid for which the students qualified (such as federal Pell Grants and state aid).¹⁰

The California Scholarship 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.¹¹ Individuals in the study were selected from participants in “Cash for College” workshops at which attendees were given assistance in 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 2 March 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 Capital 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 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. For comparison, in 2009–10 the maximum Pell Grant award was \$5,350 per year (Office of Postsecondary Education 2012). Also, the PBSs were paid directly to the students, whereas the non-PBS was paid to the institution. Table 1 shows the structure of the demonstration for the Fall 2009 cohort more specifically—the structure was similar for the Fall 2010 cohort. There are 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

9. See Richburg-Hayes et al. (2009) for more details on the programs in each site in the larger demonstration.
10. Institutions may have adjusted aid awards in response to scholarship eligibility for some students. Namely, 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/). 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), Patel and Rudd (2012), and Mayer et al. (2015).
11. 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.

Table 1. Structure of the California Program

| 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 |

Notes: 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.

Source: Ware and Patel (2012).

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 United States, 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 six credits). During spring semesters of eligibility, the entire scholarship payment was based on meeting the end-of-semester benchmark. Sixty percent of students in the program group received a performance payment for meeting the end-of-semester benchmark in their first semester after program enrollment, and 45.5 percent received an end-of-semester performance payment in their second semester (Richburg-Hayes et al. 2015).

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

Numbers of Participants

In table 2 we present information on the number of students in each cohort. In total, 5,160 individuals were recruited to be part of the PBS study; 1,720 were randomly assigned to the program-eligible group and 3,440 were assigned to the control group. We also surveyed an additional 1,500 individuals as part of our control group; they were randomly selected from the “excess” control group individuals who were not selected to be followed by MDRC for the MDRC study. Table A.1 shows means of background characteristics (at baseline) by treatment/control status. Although there are one or two characteristics that appear to differ between treatment and control groups, an omnibus

Table 2. Total (Baseline) Sample Size by Site

| Cohort | PBS | | | | | | |
|-----------|-------------------|------------|---------|---------|--------------|---------|---------|
| | Non-PBS (\$1,000) | \$500/term | | | \$1,000/term | | |
| | | 1 term | 2 terms | 4 terms | 1 term | 2 terms | 4 terms |
| Fall 2009 | 483 | 484 | 447 | 468 | 468 | 460 | |
| Fall 2010 | 653 | 637 | 679 | 611 | 633 | 637 | |
| Total | 1,136 | 1,121 | 1,126 | 1,079 | 1,101 | 1,097 | |

Note: Sample sizes include 1,500 individuals added to the control group who were not part of the MDRC study sample.

F-test yielded a *p*-value of 0.50 suggesting that randomization successfully balanced the two groups, on average.

According to Ware and Patel (2012), in the first cohort 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.

Time Use Survey¹²

To better understand the impact of PBSs 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 seven-point scale consists of five questions, such as: “When I become confused about something I’m reading, I go back and try to figure it out” (responses range from 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

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

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 and 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, both current students and those not currently enrolled were asked questions 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).¹³

We focus this analysis on time use and effort in the first semester after random assignment because the majority of both program and control students were enrolled in a postsecondary institution such that an analysis of time use is most compelling as one of the factors that determine educational success. That said, we also surveyed each cohort in the second semester after random assignment to gauge the extent to which time use changed and whether differences between the program and control group members persisted.

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.¹⁴ There were slightly more women in the study sample than in the NPSAS sample. Further, the proportion of Hispanic and black participants was much higher in this study than nationally. For example, 63 percent of participants were Hispanic compared with 15 percent nationally, and only 4 percent were black. Similarly, a language other than English is more likely to be spoken in the study sample.

4. ESTIMATION AND RESULTS

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:

$$Y_i = \alpha + \beta T_i + X_i \Theta + p_i \gamma + v_i, \quad (1)$$

13. 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 of 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.”
14. 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 sample from the NPSAS is restricted to include only students 16 to 20 years of age in order to match the age range of the study participants.

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

| Characteristics | PBS (1) | NPSAS All Types of Institutions (2) |
|---|------------|---|
| Age (years) | 17.6 | 18.4 |
| Age 17–18 (%) | 96.7 | 60.1 |
| Age 19–20 (%) | 3.2 | 39.9 |
| Female (%) | 59.9 | 53.5 |
| Race/Ethnicity (%) | | |
| Hispanic | 63.2 | 15.4 |
| Black | 3.9 | 12.3 |
| Asian | 10.8 | 5.5 |
| Native American | 0.7 | 0.9 |
| Other | 0.3 | 4 |
| First family member to attend college (%) | 54.8 | 28.7 |
| Highest degree by either parent (%) | | |
| Did not complete high school | 36.4 | 4.1 |
| High school diploma or equivalent | 30.3 | 24.6 |
| Some college including technical certificate, AA degree | 22.3 | 27.4 |
| 4-year bachelor's degree or higher | 11.1 | 43.9 |
| Non-English spoken at home | 63.0 | 12.0 |
| Number of observations | 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). We limit the NPSAS data to first-time students, enrolled at any point from 1 July through 31 December 2007. For comparability with the PBS sample in column 2, we include students aged 16–20 years who are attending any type of institution. The NPSAS means and number of observations are weighted by the 2008 study weight. AA: Associate in Arts.

where T_i is a treatment status indicator for individual i being eligible for a program scholarship, X_i is a vector of baseline characteristics (which may or may not be included), 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 nonacademic activities.¹⁵ 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 1 to obtain an effect of the treatment on each individual

15. As an alternative mechanism of grouping outcomes, we conducted a factor analysis to identify empirically determined principal components. The results roughly suggest that variables reflecting academic effort should be grouped together and those reflecting time spent on nonacademic 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.

outcome, where k refers to the k th outcome:

$$Y_k = \alpha_k + \beta_k T + \mathbf{X} \Theta_k + \boldsymbol{\rho} \gamma_k + \nu_k = \mathbf{A} \Phi_k + \nu_k. \quad (2)$$

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.

More specifically, we first estimate equation 2 (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,

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

We estimate the standard errors for β_{AVG} using the following SUR system that allows us to account for the covariance between the estimates of β_k within each domain. \mathbf{A} represents the full matrix of covariates, and Φ_k is the vector of coefficients to be estimated:

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

where I_K is a K by K identity matrix and \mathbf{A} is defined as in equation 2. 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 whereas 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).¹⁶

From our data, we focus on respondents to the survey administered late in the first semester after random assignment, although we also report some results based on second semester surveys. In the first semester, the response rate for the treatment group was 58 percent and for the control group was 40 percent; in the second semester, the response rates were 53 percent and 39 percent, respectively—differences that are statistically significant at traditional levels.¹⁷ An analysis of who responded suggests that women were more likely to respond (see table A.2). Further, respondents were somewhat more likely to be Asian and less likely to be black or Native American. They also had parents with lower education levels (more high school drop-outs and fewer high

16. For the exercise to assess the impact of survey selection on our estimates using Altonji, Elder, and Taber (2005), we normalize each outcome variable using the mean and standard deviation for the control group. We then create indexes for each domain equal to the average of the standardized values of the component outcome variables before estimation. We set the index to missing for an individual if any of the underlying outcome measures is missing. Treatment effect estimates are very similar to those estimated using SUR.

17. The overall response rate for the first semester was 45 percent, and the overall response rate for the second semester was 43 percent.

school diploma/GED holders), were less likely to speak English at home, and had somewhat higher high school GPAs. Omnibus F-tests of whether baseline characteristics jointly predict survey response status yielded a p -value of 0.000.

To create the analysis sample, we drop individuals who did not complete the time diary (or who had more than four “noncategorized” 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 for the first semester and 2,743 complete surveys for the second semester. These complete surveys represent roughly 96 percent of the total number of survey respondents in each semester. The response rates after dropping these additional observations are 56 percent and 51 percent for the treatment group in the first and second semesters, respectively. For the control group, the response rates are 39 percent and 38 percent, respectively, for the first and second semesters. Based on observable characteristics the analysis sample does not appear representative of the full experimental sample. That said, we believe these estimates are internally valid. Table A.3 shows means of background characteristics (at baseline) by treatment/control status for our analysis sample. A few characteristics appear to differ between treatment and control groups; however, omnibus F-tests are not statistically significant at conventional levels, and our estimates are largely unchanged if we include baseline characteristics as controls (see tables 4 and 5 for estimates including baseline characteristics). That said, we assess the potential impact of survey selection bias on the estimates in a separate section (see the Assessing the Possible Impact of Survey Selection Participation Bias section).

Program Impacts on Educational and Other Outcomes

In table 4 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 PBSs offered. In column 1 we provide outcome means for the control group participants. Program effect estimates with standard errors in parentheses are presented in columns 2 and 3. Note that the estimates in column 2 reflect the impact of being eligible for a PBS, and the estimates in column 3 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 4. Program effects are estimated including controls for “randomization pool” fixed effects but no other baseline characteristics.¹⁸ Corresponding estimates that also control for baseline characteristics—age; an indicator for sex (female); race/ethnicity indicators (Hispanic/Latino, black/African American, white, American Indian or Alaska native, or other); indicators for parents’ highest level of education (less than high school, high school diploma, associate’s degree); an indicator for being the first in family to attend college; indicators for speaking Spanish or English at home; and standardized responses to exit questions based on motivation—are presented in columns 5 through 7. The number of observations in each estimation sample is listed in column 8.

Focusing on the coefficients reported in column 2 of table 4, we find that PBS-eligible students were 5.2 percentage points more likely than the control group to report

18. The randomization pool fixed effects reflect the workshop region (Los Angeles, Far North, Kern County, or Capital) and cohort in which the participant was recruited.

Table 4. Estimates of PBS Impact on Academic Outcomes

| Variable | Control Mean (1) | No Baseline Controls ^a | | | Including Baseline Characteristics ^b | | | Obs (8) |
|--|---------------------|-----------------------------------|------------------|-------------------------|---|------------------|-------------------------|------------|
| | | PBS (2) | Non-PBS (3) | PBS = Non-PBS (4) | PBS (5) | Non-PBS (6) | PBS = Non-PBS (7) | |
| Ever enrolled postsecondary | 0.831 | 0.052*** (0.015) | 0.004 (0.030) | 0.124 | 0.054*** (0.015) | 0.005 (0.030) | 0.121 | 2,872 |
| Hours on all academics in last 24 hours | 4.757 | 0.277 (0.174) | 0.034 (0.345) | 0.504 | 0.262 (0.173) | 0.143 (0.344) | 0.742 | 2,874 |
| Hours studied in past 7 days | 2.936 | 0.139 (0.098) | 0.049 (0.195) | 0.659 | 0.140 (0.098) | 0.127 (0.195) | 0.946 | 2,871 |
| Prepared for last class | 0.736 | 0.073*** (0.018) | 0.027 (0.036) | 0.211 | 0.074*** (0.018) | 0.020 (0.036) | 0.155 | 2,861 |
| Attended most/all classes in past 7 days | 0.776 | 0.067*** (0.017) | 0.028 (0.033) | 0.267 | 0.070*** (0.017) | 0.031 (0.033) | 0.272 | 2,872 |
| Academic self-efficacy ^b | 0.000 | 0.121*** (0.041) | 0.024 (0.082) | 0.260 | 0.121*** (0.041) | 0.016 (0.082) | 0.225 | 2,866 |
| MSLQ index ^b | 0.000 | 0.224*** (0.042) | 0.045 (0.083) | 0.041 | 0.219*** (0.042) | 0.040 (0.084) | 0.042 | 2,871 |

Notes: Estimates obtained via ordinary least squares (OLS) regressions including location-cohort fixed effects. Time use variables refer to hours in past 24 hours unless otherwise noted. Column 4 shows the *p*-value for an F-test of the equality of the PBS and Non-PBS impacts in columns 2 and 3; column 7 shows the equivalent test for the estimates shown in columns 5 and 6. The number of observations in each estimate is shown in column 8. MSLQ: Motivated Strategies for Learning Questionnaire.

^aEstimates only include location-cohort fixed effects.

^bOutcome has been standardized using the control group distribution.

***Statistical significance at the 1% level.

ever enrolling at a postsecondary institution, a difference that is statistically significant at the 1 percent level. Further, the PBS-eligible students reported studying about eight 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 seven days, and were 6.7 percentage points more likely to report attending all or most of their classes in the last seven days. Estimates controlling for baseline characteristics (reported in column 5) are quite similar.

An important dimension to this demonstration 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 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.¹⁹

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: (1) learning strategies and (2) academic

19. A difference in impacts may also reflect a difference in the salience of the scholarships to the students, which we also discuss later.

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 4. We have standardized the variables using their respective control group means and standard deviations—the coefficients therefore reflect impacts in standard deviation units. We estimate that eligibility for a PBS had positive and statistically significant impacts on these dimensions that range from 12 to 22 percent of a standard deviation. Note as well that the impacts on learning strategies and academic self-efficacy for those 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 that there are only 24 hours in the day, a key question is what did PBS-eligible participants spend less time doing? Table 5 presents results from three other broad time-categories based on the 24-hour time diary: (1) work, (2) household production, and (3) leisure and other activities.²⁰ For work, we find only suggestive evidence that PBS-eligible participants may have spent fewer hours working. We find no evidence that the typical PBS-eligible participant spent fewer hours working during the last 24 hours and a negative but statistically insignificant reduction in hours worked in the past week. We also find no evidence that PBS eligibility reduced time spent on household production activities. Instead, we find that participants 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.

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 5 we present impacts on several potential unintended consequences. The results regarding unintended consequences are somewhat mixed. For example, on the one hand those who were eligible for a PBS were more satisfied with life and more likely to take challenging classes compared with 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 4 and 5 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

20. "Household 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.

Table 5. Estimates of PBS Impact on Quality of Nonacademic Outcomes and Potential Unintended Consequences

| Variable | Control Mean (1) | No Baseline Controls ^a | | | Including Baseline Controls ^b | | | Obs (8) |
|--|---------------------|-----------------------------------|----------------------|-------------------------|--|----------------------|-------------------------|------------|
| | | PBS (2) | Non-PBS (3) | PBS = Non-PBS (4) | PBS (5) | Non-PBS (6) | PBS = Non-PBS (7) | |
| Hours Worked in last 24 hours | 0.750 | 0.026 (0.089) | 0.101 (0.177) | 0.687 | 0.017 (0.089) | 0.057 (0.178) | 0.827 | 2,874 |
| Hours worked in past 7 days | 4.928 | -0.216 (0.399) | -0.142 (0.790) | 0.928 | -0.257 (0.399) | -0.288 (0.791) | 0.970 | 2,818 |
| Hours on household production in last 24 hours | 11.721 | 0.168 (0.147) | 0.168 (0.291) | 0.998 | 0.179 (0.147) | 0.158 (0.292) | 0.945 | 2,874 |
| Hours on leisure in last 24 hours | 6.765 | -0.482*** (0.160) | -0.302 (0.318) | 0.591 | -0.469*** (0.159) | -0.358 (0.317) | 0.738 | 2,874 |
| Times out in past 7 days | 2.077 | -0.124** (0.059) | -0.165 (0.118) | 0.746 | -0.116* (0.060) | -0.177 (0.119) | 0.624 | 2,863 |
| Strongly agree/agree to take challenging classes | 0.385 | 0.058*** (0.021) | 0.037 (0.041) | 0.638 | 0.058*** (0.021) | 0.038 (0.041) | 0.646 | 2,856 |
| Ever felt had to cheat | 0.349 | -0.106*** (0.019) | -0.106*** (0.039) | 0.989 | -0.108*** (0.020) | -0.098** (0.039) | 0.814 | 2,854 |
| External motivation ^c | 0.000 | 0.077* (0.044) | -0.132 (0.090) | 0.025 | 0.077* (0.044) | -0.104 (0.089) | 0.052 | 2,417 |
| Internal motivation ^c | 0.000 | 0.019 (0.045) | -0.124 (0.092) | 0.136 | 0.017 (0.045) | -0.124 (0.091) | 0.137 | 2,419 |
| Ever asked for regrade | 0.197 | 0.006 (0.017) | -0.087*** (0.033) | 0.008 | 0.004 (0.017) | -0.087*** (0.034) | 0.010 | 2,860 |
| Very satisfied/satisfied with life | 0.624 | 0.010 (0.020) | -0.002 (0.041) | 0.781 | 0.011 (0.020) | -0.003 (0.041) | 0.741 | 2,850 |

Notes: Estimates obtained via OLS regressions including location-cohort fixed effects. Column 4 shows the *p*-value for an F-test of the equality of the PBS and non-PBS impacts in columns 2 and 3; column 7 shows the equivalent test for the estimates shown in columns 5 and 6. The number of observations in each estimate is shown in column 8.

^aEstimates only include location-cohort fixed effects.

^bEstimates include controls location-cohort fixed effects and baseline characteristics: age; an indicator for sex is female; race/ethnicity indicators for Hispanic/Latino, black/African American, white, American Indian or Alaska native, or other; indicators for parents' highest level of education is less than high school, high school diploma, associate's degree; an indicator for being the first in family to attend college; indicators for speaking Spanish or English at home; and standardized responses to exit questions based on motivation.

^cThe index has been standardized using the control group distribution.

*Statistical significance at the 10% level; **statistical significance at the 5% level; ***statistical significance at the 1% level.

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 6 we combine the individual outcomes into four “domains” using the SUR approach described above. Specifically, we focus on academic activities, quality of educational input, nonacademic activities, and unintended consequences.²¹ The impacts reported in table 6 and subsequent tables have been standardized such that they represent average impacts as a percentage of the control group

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.” “Nonacademic 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 regrade,” and “Very satisfied/satisfied with life.” We do not include whether an individual had “ever enrolled” in a postsecondary institution in the “all academic activities” index as it represents an academic decision on the extensive margin rather than the intensive margin.

Table 6. Index Estimates of PBS Impact

| | Full Sample | | | | Conditional on Enrolling in a Postsecondary Institution | | | |
|--------------------------------------|----------------------|-----------------------|----------------------------------|------------|---|-----------------------|----------------------------------|------------|
| | PBS Impact (1) | Non-PBS Impact (2) | p-value for PBS = Non-PBS (3) | Obs (4) | PBS Impact (5) | Non-PBS Impact (6) | p-value for PBS = Non-PBS (7) | Obs (8) |
| All academic activities | 0.113*** (0.027) | 0.039 (0.056) | 0.203 | 2874 | 0.052** (0.026) | 0.060 (0.050) | 0.876 | 1,840 |
| Quality of educational input | 0.173*** (0.035) | 0.034 (0.075) | 0.077* | 2872 | 0.161*** (0.040) | 0.015 (0.085) | 0.091* | 1,840 |
| Nonacademic activities | -0.035* (0.018) | -0.023 (0.034) | 0.730 | 2874 | -0.021 (0.021) | -0.028 (0.039) | 0.862 | 1,840 |
| Unintended consequences ^a | -0.048*** (0.018) | -0.087** (0.037) | 0.315 | 2874 | -0.038* (0.021) | -0.085** (0.043) | 0.282 | 1,840 |

Notes: Estimates are indexed estimates obtained via the seemingly unrelated regression (SUR) strategy discussed in the paper. All regressions include location-cohort fixed effects. Column 3 shows the p-value for an F-test of the equality of the PBS and non-PBS impacts in columns 1 and 2; column 7 shows the equivalent test for the estimates shown in columns 5 and 6. The number of observations in each estimation sample is shown in columns 4 and 8.

^aIn constructing the index, components are adjusted so that a negative indicates a “good” outcome.

*Statistical significance at the 10% level; **statistical significance at the 5% level; ***statistical significance at the 1% level.

standard deviation. Note that now we estimate a positive impact on academic activities of about 10 percent of a standard deviation and that the impact is 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 nonacademic activities, although the coefficient estimate is only significant at the 10 percent level. Further, 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 time allocation in the sense that students spend more time and effort on academic activities and less time on other activities.

To explore the possibility that results are driven by an incentive effect only on the extensive margin, we have reestimated the impacts limiting the samples to those students who enrolled in school, as shown in columns 5 through 8 of table 6. We find that estimated impacts on academic activities (including quality of academic input) 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. Further, we note that the results are based on self-reported data and therefore the possibility that respondents’ self-reports are correlated with their treatment status may also bias the estimated impacts.

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

Three key questions are (1) whether the size of the potential scholarship affects the impact on student behavior, (2) whether scholarship eligibility impacts student behavior even after incentives are removed, and (3) whether it is the incentive structure or the additional income that generates changes in student behavior. Prior studies of PBSs have tended to focus on one type of scholarship of a particular duration with variation

Table 7. PBS Impact in the First Semester by Scholarship Size and Incentive

| | Scholarship Size and Incentive | | | | | Obs |
|---|--------------------------------|---------------------|---------------------|-------------------------------------|----------------------------------|-------|
| | \$500/Term (1) | \$1,000/Term (2) | Non-PBS (3) | \$500/Term = \$1,000/Term (4) | Non-PBS = \$1,000/Term (5) | |
| Ever enrolled ^a | 0.072*** (0.021) | 0.039** (0.018) | 0.004 (0.030) | 0.188 | 0.295 | 2,872 |
| Currently enrolled ^a | 0.071*** (0.023) | 0.049** (0.020) | 0.028 (0.032) | 0.430 | 0.554 | 2,873 |
| All academic activities ^b | 0.120*** (0.036) | 0.108*** (0.033) | 0.039 (0.056) | 0.772 | 0.260 | 2,874 |
| Quality of educational input ^b | 0.191*** (0.048) | 0.160*** (0.043) | 0.034 (0.075) | 0.597 | 0.125 | 2,872 |
| Nonacademic activities ^b | -0.026 (0.025) | -0.041* (0.022) | -0.023 (0.034) | 0.630 | 0.632 | 2,874 |
| Unintended consequences ^{b,c} | -0.055** (0.026) | -0.044* (0.022) | -0.087** (0.037) | 0.731 | 0.288 | 2,874 |

Notes: Column 5 shows the *p*-value for an F-test of the equality of the PBS and non-PBS impacts.

^aEstimates obtained via OLS regressions.

^bEstimates obtained via the SUR strategy discussed in the paper.

^cIn constructing the index, components are adjusted so that a negative indicates a "good" outcome.

*Statistical significance at the 10% level; **statistical significance at the 5% level; ***statistical significance at the 1% level.

in other types of resources available to students (such as student support services or a counselor). In this study, students eligible for a PBS were also randomly assigned to scholarships of differing durations and/or sizes as well as a non-PBS. As noted in table 1, 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 to study the impact of scholarship characteristics on student behavior.²² We present results using SUR in tables 7 and 8.²³

Impacts by Size of the Scholarship

Theoretically one would expect that the larger-sized scholarships would induce increased effort compared with 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 difference between scholarships worth \$1,000 and those worth \$500 per term. In table 7, 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. The *p*-values for the test of equality of the coefficient estimates in columns 1 and 2 are

22. Although 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.

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

Table 8. PBS Impact in the Second Semester by Scholarship Length and Incentive

| | Scholarship Length | | | | | Obs (6) |
|---|--------------------|----------------------|----------------------|----------------------------|----------------------------|------------|
| | 1 Term (1) | 2+ Terms (2) | Non-PBS (3) | 1 Term = 2 Terms (4) | Non-PBS = 1 Term (5) | |
| Ever enrolled ^a | 0.032 (0.028) | 0.013 (0.015) | 0.003 (0.029) | 0.534 | 0.470 | 2,742 |
| Currently enrolled ^a | 0.003 (0.032) | 0.016 (0.017) | 0.015 (0.033) | 0.704 | 0.795 | 2,740 |
| All academic activities ^b | 0.038 (0.057) | 0.082*** (0.030) | -0.017 (0.057) | 0.468 | 0.477 | 2,743 |
| Quality of educational input ^b | 0.139* (0.078) | 0.125*** (0.038) | 0.108 (0.076) | 0.861 | 0.766 | 2,742 |
| Nonacademic activities ^b | 0.029 (0.039) | -0.066*** (0.019) | -0.007 (0.034) | 0.021 | 0.470 | 2,743 |
| Unintended consequences ^{b,c} | -0.063* (0.037) | -0.067*** (0.019) | -0.104*** (0.036) | 0.923 | 0.420 | 2,742 |

Notes: Regressions include cohort-location fixed effects. Column 4 shows the *p*-value for an F-test of the equality of the PBS 1 term and PBS 2 or more terms impacts. Column 5 shows the *p*-value for an F-test of the equality of the PBS and non-PBS impacts.

^aEstimates obtained via OLS regressions.

^bEstimates obtained via the SUR strategy discussed in the paper.

^cIn constructing the index, components are adjusted so that a negative indicates a "good" outcome.

*Statistical significance at the 10% level; ***statistical significance at the 1% level.

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. Although 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 2017). This finding remains a puzzle, although we offer some potential explanations for further consideration in section 5.

Impacts by Duration of the Scholarship

If incentives have no effect on the GPA production function, 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 (Benabou and Tirole 2003; Huffman and Bognanno, forthcoming). Alternatively, one might expect increased effort in the first semester to be habit-forming or make students more efficient at transforming study time into a higher 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 table 8 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 nonacademic activities and unintended consequences. As such, we find that a PBS 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 PBSs 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 PBS compared with just awarding additional money. This test can be made in table 7 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.²⁴ 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

24. 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.

was paid to the institutions. As such, the non-PBS may not have been as salient to the student. There is also the possibility that institutions at least partly offset the non-PBS with reductions in other forms of financial aid. One might interpret the fact that the PBS (column 2 of table 7) had a larger impact on enrollment than the non-PBS (column 3 of table 7) as providing support for these alternative interpretations. If the PBSs and non-PBSs were awarded in the same way, 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 suggests the students may have been unaware of the award or that institutions offset the award with reductions in other financial aid.

Although 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 offset the non-PBS with reductions in other financial aid, which is unlikely since 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 a \$1,000 subsidy increases enrollment by about 4 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 PBSs on enrollment, when scaled by how much students actually received, is larger.

Taken together, we believe the evidence suggests that incentives played a key role in changing student behavior. We note this finding is also consistent with Scott-Clayton (2011), who studies the West Virginia PROMISE scholarship—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.

Assessing the Possible Impact of Survey Selection Participation Bias

As we noted earlier, the overall survey response rate was only about 58 percent, and response rates for PBS-eligible participants were statistically greater than response rates for students in the control group. We explore the extent to which selection may be affecting our PBS impact estimates in table 9. We reproduce the table 6 estimates in column 1 for ease of comparison.

In column 2 of table 9 we simply include controls for baseline characteristics—student age, sex, race/ethnicity, parental education, first in family to attend college, English or Spanish spoken at home, and standardized responses to exit questions on motivation.²⁵ In column 3 we use the baseline controls to predict treatment status and

25. The motivation questions asked participants to respond on a Likert-type scale of agreement to the following: (1) If I complete a FAFSA right away it's because I want people to think I'm ready for college; (2) I follow advice on how to pay for college because it will help me go to college; (3) If I complete a FAFSA it's because I want to get as much financial aid as possible to go to college; and (4) If I complete a FAFSA it's because it makes me happy to be closer to my goal of going to college.

Table 9. Index Estimates of the PBS Impacts: Exploring the Role of Selection

| | PBS Impact ^a | PBS impact Including Baseline Controls ^a | IPW Estimate of PBS Impact ^a | Estimate of Bias from Selection ^b | Implied Ratio of Selection on Unobservables to Selection on Observables |
|--------------------------------------|-------------------------|---|---|--|---|
| | (1) | (2) | (3) | (4) | (5) |
| All academic activities | 0.113*** (0.027) | 0.114*** (0.027) | 0.110*** (0.027) | -0.042 (0.140) | -2.648 |
| Quality of educational input | 0.173*** (0.035) | 0.170*** (0.035) | 0.164*** (0.035) | 0.256 (0.273) | 0.653 |
| Nonacademic activities | -0.035* (0.018) | -0.034* (0.018) | -0.036** (0.018) | 0.140 (0.064) | -0.271 |
| Unintended consequences ^c | -0.048*** (0.018) | -0.049*** (0.018) | -0.048** (0.019) | 0.164 (0.128) | -0.288 |

Notes: Column 1 estimates simply replicate the index estimates presented in column 1 of table 6. The estimates in column 2 additionally control for baseline characteristics. Estimates in column 3 use inverse-probability weighting to adjust for selection in the underlying regressions. The weights are calculated using baseline characteristics: age; an indicator for sex is female; race/ethnicity indicators for Hispanic/Latino, black/African American, white, American Indian or Alaska native, or other; indicators for parents' highest level of education is less than high school, high school diploma, associate's degree; an indicator for being the first in family to attend college; indicators for speaking Spanish or English at home; and standardized responses to exit questions based on motivation. Column 4 is the estimate of the bias in the treatment estimate using Altonji, Elder, and Taber (2005) under the Condition 4 assumption that the standardized selection on unobservables is equal to the standardized selection on observables. The implied ratio in column 5 is the ratio of the standardized selection on unobservables to the standardized selection on observables that is consistent with no treatment effect. All outcome regressions control for location-cohort fixed effects. IPW: inverse probability weighting.

^aEstimates use the SUR strategy discussed in the paper.

^bEstimates obtained by constructing the index outcomes before estimation and dropping observations receiving the non-PBS treatment.

^cIn constructing the index, components are adjusted so that a negative indicates a "good" outcome.

*Statistical significance at the 10% level; **statistical significance at the 5% level; ***statistical significance at the 1% level.

then use inverse probability weighting to adjust the impact estimates for selection. Adjusting for selection either by controlling for baseline characteristics directly or by inverse probability weighting has very little effect on our estimates of the impact of PBS eligibility on any of the index outcomes. Although both strategies can only address the potential for selection on observable characteristics, they suggest that selection on unobservable characteristics would have to be quite large to overturn our estimates.

In the final columns of table 9, we make use of the strategy of Altonji, Elder, and Taber (2005) to estimate the bias in the PBS estimate under the assumption of equal (standardized) selection on observables and unobservables (column 4) and to calculate the implied ratio of selection on unobservables to selection on observables under the assumption of no PBS effect (column 5).²⁶ For the index of all academic activities, the index of nonacademic activities, and the index of unintended consequences, the estimates assuming equal selection on observables and unobservables suggest that the bias is pushing the treatment effect estimates toward zero. In other words, in the absence of selection bias we would estimate larger treatment effects in absolute value. For the index of quality of educational input, the estimate of the bias assuming equal selection

26. For this exercise, we drop individuals assigned to the non-PBS treatment group. This does not affect the PBS estimates.

on observables and unobservables suggests that there may be no positive impact of PBS eligibility on the quality of educational input.

So how much selection on unobservables is necessary to overturn our findings? When we translate the bias estimates into ratios of selection on unobservables to selection on observables consistent with no treatment impact, we find that selection on unobservables would have to be 2.6 times as large as selection on observables and in the opposite direction in order to conclude that there is no impact of PBS eligibility on the index of all academic activities. In other words, if survey nonresponse generates positive selection into treatment status based on observable characteristics, the selection must be negative (and much larger) based on unobservable characteristics. This seems unlikely and therefore we conclude that PBS eligibility likely has a positive effect on academic activities. Similarly, we believe that selection on unobservables is unlikely to explain the estimates of the impact of PBS eligibility on nonacademic activities and unintended consequences. In both cases the degree of selection on unobservables does not have to be very large relative to the selection on observables but it does have to be in the opposite direction. In contrast, for the quality of educational inputs, selection on unobservables is more likely to be able to explain our estimated treatment effect. Selection on unobservables that is roughly 65 percent as large as selection on observables would explain all the estimated effect of PBS eligibility on quality of educational inputs.

5. CONCLUSION

Although education policy makers have become increasingly interested in using incentives to improve educational outcomes at the postsecondary level, the evidence continues to generate mixed impacts, leading to the question of whether such incentives can actually change student effort toward 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 PBSs 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 PBS, suggesting such incentives do not permanently change students' cost of effort or their ability to transform effort into educational outcomes.

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. Finally, these results could also be consistent with students not fully understanding their own "education production function," that is, how their own effort and ability will be transformed into academic outcomes like grades. Although 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.²⁷ Our data do not allow us to thoroughly understand why larger incentive payments did not generate larger changes in behavior, yet understanding why they did not would be important for the optimal design of incentive schemes to improve educational attainment.

Finally, while the evidence from this study of PBSs 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.

ACKNOWLEDGMENTS

We thank Eric Auerbach, Elijah de la Campa, Ross Cole, 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 Reshma Patel for extensive help in understanding the MDRC data. We are also grateful to Todd Elder for sharing programs for exploring selection bias. Orley Ashenfelter, Todd Elder, Alan Krueger, Jonas Fisher, Luoqia Hu, Derek Neal, Reshma Patel, Lashawn Richburg-Hayes, and Shyam Gouri Suresh, 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. The 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.

REFERENCES

- Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113(1): 151–184. doi:10.1086/426036.
- 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): 136–163. doi:10.1257/app.1.1.136.
- Angrist, Joshua, and Victor Lavy. 2009. The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American Economic Review* 99(4): 1384–1414. doi:10.1257/aer.99.4.1384.
- Angrist, Joshua, Philip Oreopoulos, and Tyler Williams. 2014. When opportunity knocks, who answers? New evidence on college achievement awards. *Journal of Human Resources* 49(3): 572–610.
- Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar. 2009. Large stakes and big mistakes. *Review of Economic Studies* 76(2): 451–469. doi:10.1111/j.1467-937X.2009.00534.x.

27. Although 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.

Babcock, Philip, and Mindy Marks. 2011. The falling time cost of college: Evidence from half a century of time use data. *Review of Economics and Statistics* 93(2): 468–478. doi:10.1162/REST_a_00093.

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* 32(3): 563–599. doi:10.1086/675229.

Barrow, Lisa, and Cecilia Elena Rouse. 2013. Documentation for the Princeton University PBS time use survey. Unpublished paper, Princeton University.

Becker, Gary S. 1967. *Human capital and the personal distribution of income*. Ann Arbor: University of Michigan Press.

Benabou, Roland, and Jean Tirole. 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies* 70(3): 489–520. doi:10.1111/1467-937X.00253.

Casey, Marcus, Jeffrey Cline, Ben Ost, and Javaeria Quereshi. 2018. Academic probation, student performance and strategic course-taking. *Economic Inquiry* 56(3): 1646–1677. doi:10.1111/ecin.12566.

Cha, Paulette, and Reshma Patel. 2010. *Rewarding progress, reducing debt: Early results from the performance-based scholarship demonstration in Ohio*. New York: MDRC.

Cornwell, Christopher M., Kyung Hee Lee, and David B. Mustard. 2005. Student responses to merit scholarship rules. *Journal of Human Resources* 50(4): 895–917. doi:10.3368/jhr.XL.4.895.

Das, Jishnu, Quy-Toan Do, and Berk Ozler. 2005. Reassessing conditional cash transfer programs. *World Bank Research Observer* 20(1): 57–80. doi:10.1093/wbro/lki005.

Deci, Edward L. 1975. *Intrinsic motivation*. New York: Plenum. doi:10.1007/978-1-4613-4446-9.

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(6): 627–668. doi:10.1037/0033-2909.125.6.627.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic motivation and self-determination in human behavior*. New York: Plenum. doi:10.1007/978-1-4899-2271-7.

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, pp. 208–302. Chicago: University of Chicago Press.

Dynarski, Susan M. 2008. Building the stock of college-educated labor. *Journal of Human Resources* 43(3): 576–610.

Fryer, Roland G., Jr. 2011. Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics* 126(4): 1755–1798. doi:10.1093/qje/qjr045.

Gneezy, Uri, and Aldo Rustichini. 2000. Pay enough or don't pay at all. *Quarterly Journal of Economics* 115(3): 791–810. doi:10.1162/003355300554917.

Huffman, David, and Michael Bognanno. Forthcoming. High-powered performance pay and crowding out of nonmonetary motives. *Management Science*. doi:10.1287/mnsc.2017.2846.

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. The effects of an 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(4): 442–453. doi:10.1086/269336.

Kelly, Dana, Holly Xie, Christine Winquist Nord, Frank Jenkins, Jessica Ying Chan, and David Kastberg. 2013. *Performance of U.S. 15-year-old students in mathematics, science, and reading literacy in an international context: First look at PISA 2012*. Washington, DC: U.S. Department of Education (NCES 2014–024).

Kling, Jeffrey R., and Jeffrey B. Liebman. 2004. Experimental analysis of neighborhood effects on youth. Princeton University Industrial Relations Section Working Paper No. 483.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. Incentives to learn. *Review of Economics and Statistics* 91(3): 437–456. doi:10.1162/rest.91.3.437.

Lau, Yan. 2017. Tournament incentives structure and effort, or the art of carrot dangling. Unpublished paper, Reed College.

Lee, David S. 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies* 76(3): 1071–1102. doi:10.1111/j.1467-937X.2009.00536.x.

Lindo, Jason M., Nicholas J. Sanders, and Philip Oreopoulos. 2010. Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics* 2(2): 95–117. doi:10.1257/app.2.2.95.

Mayer, Alexander, Reshma Patel, Timothy Rudd, and Alyssa Ratledge. 2015. *Designing scholarships to improve college success: Final report on the performance-based scholarship demonstration*. New York: MDRC.

Midgley, Carol, Martin L. Maehr, Ludmila Z. Hruda, Eric Anderman, Lynley Anderman, Kimberley E. Freeman, Margaret Gheen, et al. 2000. *Manual for the patterns of adaptive learning scales*. Ann Arbor: University of Michigan Press.

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.

Office of Postsecondary Education. 2012. *2010–2011 Federal Pell Grant end-of-the-year report*. Washington, DC: U.S. Department of Education.

Organisation for Economic Co-operation and Development (OECD). 2011. How many students finish tertiary education? In *Education at a glance 2011: OECD indicators*, edited by Angel Gurria, pp. 60–71. Paris: OECD Publishing. doi:10.1787/eag-2011-7-en.

Patel, Reshma, Lashawn Richburg-Hayes, Elijah de la Campa, and Timothy Rudd. 2013. *Performance-based scholarships: What have we learned? Interim findings from the PBS demonstration*. New York: MDRC.

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. doi:10.1037/0022-0663.82.1.33.

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)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning Technical Report No. 91-B-004.

Provasnik, Stephen, David Kastberg, David Ferraro, Nita Lemanski, Stephen Roey, and Frank Jenkins. 2012. *Highlights from TIMSS 2011: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Rawlings, Laura, and Gloria Rubio. 2005. Evaluating the impact of conditional cash transfer programs. *World Bank Research Observer* 20(1): 29–55. doi:10.1093/wbro/lki001.

Richburg-Hayes, Lashawn, Paulette Cha, Monica Cuevas, Amanda Grossman, Reshma Patel, and Colleen Sommo. 2009. *Paying for college success: An introduction to the performance-based scholarship demonstration*. New York: MDRC.

Richburg-Hayes, Lashawn, Reshma Patel, Thomas Brock, Elijah de la Campa, Timothy Rudd, and Ireri Valenzuela. 2015. *Providing more cash for college: Interim findings from the performance-based scholarship demonstration in California*. New York: MDRC.

Romer, David. 1993. Do students go to class? Should they? *Journal of Economic Perspectives* 7(3): 167–174. doi:10.1257/jep.7.3.167.

Schudde, Lauren, and Judith Scott-Clayton. 2016. Pell grants as performance-based scholarships? An examination of satisfactory academic progress requirements in the nation's largest need-based aid program. *Research in Higher Education* 57(8): 943–967. doi:10.1007/s11162-016-9413-3.

Scott-Clayton, Judith. 2011. On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources* 46(3): 614–646.

Sjoquist, David L., and John V. Winters. 2012. Building the stock of college-educated labor revisited. *Journal of Human Resources* 47(1): 270–285.

Stinebrickner, Todd R., and Ralph Stinebrickner. 2008. The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy* 8(1). doi:10.2202/1935-1682.1868.

U.S. Commission on Excellence in Education, Department of Education. 1983. *A nation at risk: The imperative for educational reform. A report to the Nation and the Secretary of Education, United States Department of Education*. Washington, DC: The Commission.

Ware, Michelle, and Reshma Patel. 2012. *Does more money matter? An introduction to the performance-based scholarship demonstration in California*. New York: MDRC.

APPENDIX

Table A.1. Randomization of Program and Control Groups

| Baseline Characteristic (%) | Random Assignment | | <i>p</i> -value of Difference | <i>N</i> |
|---------------------------------------|-------------------|---------------|-------------------------------|----------|
| | Program Group | Control Group | | |
| Age (years) | 17.6 | 17.6 | 0.251 | 6,660 |
| Female | 60.6 | 59.7 | 0.526 | 6,659 |
| Race/Ethnicity ^a | | | | |
| Hispanic | 63.1 | 63.2 | 0.895 | 6,597 |
| Black | 3.3 | 4.1 | 0.173 | 6,597 |
| White | 18.2 | 18.7 | 0.562 | 6,597 |
| Asian | 10.6 | 10.8 | 0.817 | 6,597 |
| Native American | 0.7 | 0.7 | 0.784 | 6,597 |
| Other | 0.4 | 0.3 | 0.534 | 6,597 |
| Multi-racial | 3.6 | 2.2 | 0.001 | 6,597 |
| Race not reported | 0.8 | 1.0 | 0.507 | 6,660 |
| Highest degree by either parent | | | | |
| No high school diploma | 36.6 | 36.3 | 0.828 | 6,541 |
| High school diploma/GED | 29.8 | 30.5 | 0.596 | 6,541 |
| Associate's or similar degree | 23.0 | 22.1 | 0.430 | 6,541 |
| Bachelor's degree | 10.7 | 11.2 | 0.561 | 6,541 |
| First family member to attend college | 56.4 | 54.8 | 0.249 | 6,612 |
| Primary language | | | | |
| English | 37.9 | 36.6 | 0.248 | 6,617 |
| Spanish | 50.7 | 51.5 | 0.512 | 6,617 |
| Other language | 11.4 | 11.9 | 0.563 | 6,617 |
| High school grade point average | 2.91 | 2.90 | 0.804 | 4,890 |

Notes: Calculations using Baseline Information Form (BIF) data for all experiment participants (6,660 individuals) whom we attempted to survey at the end of their first program semester. For cohort 1, this is Fall 2009. For cohort 2 this is Fall 2010. The means have been adjusted by research cohort and workshop region. High school GPA data were not collected by MDRC for the 1,500 control group individuals not included in the MDRC study sample. An omnibus F-test of whether baseline characteristics jointly predict program group status (including baseline questions regarding motivation not shown in this table but excluding high school GPA) yields a *p*-value of 0.498. (The *p*-value equals 0.668 using the smaller sample for which high school GPA data are available.) Distributions may not add to 100 percent because of rounding.

^aRespondents 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.

Table A.2. Representativeness of the Analysis Sample

| Baseline Characteristic (%) | Analysis Sample | Nonrespondents | <i>p</i> -value of Difference | <i>N</i> |
|-----------------------------|-----------------|----------------|-------------------------------|----------|
| Age (years) | 17.6 | 17.6 | 0.072 | 6,660 |
| Female | 63.6 | 57.1 | 0.000 | 6,659 |
| Race/Ethnicity ^a | | | | |
| Hispanic | 62.2 | 63.9 | 0.105 | 6,597 |
| Black | 3.3 | 4.4 | 0.028 | 6,597 |
| White | 18.7 | 18.5 | 0.760 | 6,597 |
| Asian | 12.4 | 9.5 | 0.000 | 6,597 |
| Native American | 0.4 | 0.9 | 0.029 | 6,597 |
| Other | 0.3 | 0.3 | 0.903 | 6,597 |

Table A.2. Continued.

| Baseline Characteristic (%) | Analysis Sample | Nonrespondents | <i>p</i> -value of Difference | <i>N</i> |
|---------------------------------------|-----------------|----------------|-------------------------------|----------|
| Multi-racial | 2.6 | 2.5 | 0.753 | 6,597 |
| Race not reported | 0.9 | 1.0 | 0.768 | 6,660 |
| Highest degree by either parent | | | | |
| No high school diploma | 37.7 | 35.3 | 0.041 | 6,541 |
| High school diploma/GED | 28.6 | 31.6 | 0.008 | 6,541 |
| Associate's or similar degree | 22.4 | 22.2 | 0.879 | 6,541 |
| Bachelor's degree | 11.4 | 10.8 | 0.497 | 6,541 |
| First family member to attend college | 54.8 | 55.5 | 0.568 | 6,612 |
| Primary language | | | | |
| English | 34.9 | 38.5 | 0.000 | 6,617 |
| Spanish | 51.8 | 50.9 | 0.429 | 6,617 |
| Other language | 13.3 | 10.6 | 0.001 | 6,617 |
| High School GPA | 3.0 | 2.8 | 0.000 | 4,890 |

Notes: Calculations using Baseline Information Form (BIF) data for all experiment participants (6,660 individuals) whom we attempted to survey at the end of their first program semester. For cohort 1, this is Fall 2009. For cohort 2 this is Fall 2010. High school GPA data were not collected by MDRC for the 1,500 control group individuals not included in the MDRC study sample. An omnibus F-test of whether baseline characteristics jointly predict analysis group status yielded a *p*-value of 0.000.

^aRespondents 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.

Table A.3. Randomization of Program and Control Groups, Analysis Sample

| Baseline Characteristic (%) | Random Assignment | | <i>p</i> -value of Difference | <i>N</i> |
|---------------------------------------|-------------------|---------------|-------------------------------|----------|
| | Program Group | Control Group | | |
| Age (years) | 17.6 | 17.6 | 0.109 | 2,874 |
| Female | 63.0 | 63.8 | 0.687 | 2,874 |
| Race/Ethnicity ^a | | | | |
| Hispanic | 62.1 | 62.1 | 0.988 | 2,847 |
| Black | 3.2 | 3.5 | 0.732 | 2,847 |
| White | 18.2 | 18.7 | 0.623 | 2,847 |
| Asian | 11.8 | 12.9 | 0.367 | 2,847 |
| Native American | 0.5 | 0.4 | 0.673 | 2,847 |
| Other | 0.6 | 0.2 | 0.033 | 2,847 |
| Multi-racial | 3.6 | 2.2 | 0.030 | 2,847 |
| Race not reported | 0.8 | 1.0 | 0.559 | 2,874 |
| Highest degree by either parent | | | | |
| No high school diploma | 37.4 | 38.2 | 0.670 | 2,837 |
| High school diploma/GED | 28.4 | 28.8 | 0.822 | 2,837 |
| Associate's or similar degree | 24.0 | 21.3 | 0.100 | 2,837 |
| Bachelor's degree | 10.3 | 11.8 | 0.228 | 2,837 |
| First family member to attend college | 54.4 | 55.0 | 0.751 | 2,860 |

Table A.3. Continued.

| Baseline Characteristic (%) | Random Assignment | | <i>p</i> -value of Difference | <i>N</i> |
|---------------------------------|-------------------|---------------|-------------------------------|----------|
| | Program Group | Control Group | | |
| Primary language | | | | |
| English | 35.6 | 34.4 | 0.435 | 2,860 |
| Spanish | 51.5 | 51.9 | 0.846 | 2,860 |
| Other language | 12.9 | 13.8 | 0.499 | 2,860 |
| High school grade point average | 2.97 | 3.02 | 0.037 | 2,165 |

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 being in the program group (including baseline questions regarding motivation not shown in this table but excluding high school GPA) yields a *p*-value of 0.328. (The *p*-value equals 0.117 using the smaller sample for which high school GPA data are available.) Distributions may not add to 100 percent because of rounding.

^a 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.