When Opportunity Knocks, Who Answers?

New Evidence on College Achievement Awards

Joshua Angrist
Philip Oreopoulos
Tyler Williams

ABSTRACT

We evaluate the effects of academic achievement awards for first- and second-year college students studying at a Canadian commuter college. The award scheme offered linear cash incentives for course grades above 70. Awards were paid every term. Program participants also had access to peer advising by upperclassmen. Program engagement appears to have been high but overall treatment effects were small. The intervention increased the number of courses graded above 70 and points earned above 70 for second-year students but generated no significant effect on overall GPA. Results are somewhat stronger for a subsample of applicants who correctly described the program rules.

I. Introduction

As college enrollment rates have increased, so too have concerns about rates of college completion. Around 45 percent of American college students and nearly 25 percent of Canadian college students fail to complete any college program within six years of postsecondary enrollment (Shaienks and Gluszynksi 2007; Shapiro et al. 2012). Those who do finish take much longer than they used to (Turner 2004; Bound, Lovenheim, and Turner 2010; Babcock and Marks 2011). Delays and dropouts may be both privately and socially costly. Struggling college students often

Joshua Angrist is a professor of economics at MIT. Philip Oreopoulos is a professor of economics at the University of Toronto. Tyler Williams is an economist at Amazon.com, Inc. The data used in this article can be obtained beginning January 2015 through December 2017 from Joshua Angrist, 50 Memorial Drive, Building E52, Room 353, Cambridge MA 02142-1347, angrist@mit.edu. *The authors thank the Higher Education Quality Council of Ontario and the Spencer Foundation for funding this work. They also thank participants in the MIT Labor Lunch, the Harvard Labor Economics Workshop, and the MIT Labor/Public Finance Workshop for helpful comments.*
[Submitted August 2012; accepted March 2013]
ISSN 0022-166X E-ISSN 1548-8004

show little evidence of skill improvement (Arum and Roksa 2011). They pay a higher cost in foregone earnings than those who finish while losing the benefit of any possible "sheepskin effects" from degree completion. Time on campus is also subsidized at public colleges and universities so repeated course failures and long completion times are costly for taxpayers. A recent analysis by Harris and Goldrick-Rab (2010) shows steadily declining degree-to-expenditure ratios in American public colleges, a trend generated by falling completion rates as well as increasing sticker prices.

In an effort to boost grades and on-time graduation rates, most universities deploy an array of support services. These efforts reflect a practical response to an important problem, but evidence that academic support services improve outcomes is mixed at best. A randomized trial discussed by Scrivener and Weiss (2009) finds that campus support services generate small improvements in grades and reduce student attrition, but Angrist, Lang, and Oreopoulos (2009) and MacDonald et al. (2009) find little effect from support services. Part of the problem seems to be that takeup rates for most support services are low. More proactive programs that facilitate higher takeup and more intensive support appear more successful than relatively passive interventions offering only "service availability" (Scrivener, Sommo, and Collado 2009; Bettinger and Baker 2011).

A parallel effort looks to boost college achievement and completion with financial incentives. Traditional need-based grant aid—which makes up the bulk of North American aid— flows to recipients in a manner that is mostly independent of academic performance while embedding little incentive for timely degree completion. Merit-based aid, on the other hand, depends on academic achievement. Most merit awards go to top-performing students, who can be expected to do reasonably well with or without support. Performance-based awards for students not already on top are a new but rapidly expanding policy development. If successful, such awards may improve academic outcomes, increase the rate of degree completion, and ultimately save both taxpayers and recipients money.

Georgia's Helping Outstanding Pupils Educationally (HOPE) program, introduced in 1993, is a pioneering effort in this direction. Funded by lottery ticket sales, HOPE covers tuition and fees at any public college or university for students who earned at least a 3.0 high school GPA. Students lose the HOPE scholarship if their college GPA dips below 3.0. Georgia HOPE has been a model for dozens of similar state programs. Accumulating empirical evidence suggests HOPE-like award schemes improve high school achievement. (See, for example, Pallais 2009.) On the other hand, such programs also appear to reduce recipients' college course loads (Cornwell, Lee, and Mustard 2005), increase their consumption (Cornwell and Mustard 2007), and reduce attendance at out-of-state colleges and college quality (Cornwell, Mustard, and Sridhar 2006; Cohodes and Goodman forthcoming).

Estimates of the effects of HOPE-style programs on college enrollment and completion are mixed. Dynarski (2008) reports large increases in Georgia and Arkansas's college-educated populations a few years after the introduction of HOPE and a similar Arkansas program, while Castleman (2014) estimates that Florida's HOPE-style public university scholarship boosted recipients' in-state public college completion rates. By contrast, recent analyses by Sjoquist and Winters (2012a, 2012b) find no effect when looking at a broader range of state programs with more recent data and updated clustered standard error estimation.

Most research on HOPE-style programs uses observational designs. Among the most credible of these evaluations, Scott-Clayton's (2011) regression discontinuity investigation of West Virginia's Providing Real Opportunities for Maximizing In-State Student Excellence (PROMISE) scholarship generates evidence of substantial increases in four- and five-year graduation rates. Importantly, however, this study shows the PROMISE scholarship increased GPAs and credits earned during the first three years of college only, when students faced a minimum GPA requirement to maintain award eligibility. This suggests that the incentive effects of the scholarships are larger than the income effects resulting from greater financial aid.

Incentive experiments and quasi-experimental research designs in European universities have also produced mixed results. Using a regression-discontinuity design, Garibaldi et al. (2012) found that higher tuition induces faster degree completion by Italian women. De Paola, Scoppa, and Nistico (2012) also find substantial positive effects of a randomized financial award for business administration students in southern Italy. On the other hand, randomized evaluations of financial incentives offered to Dutch university students generated little overall effect (Leuven, Oosterbeek, and van der Klaauw 2010; Leuven, et al. 2011).[1]

In an effort to encourage on-time completion and retention, a few incentive programs target college credits for those already enrolled. In a randomized evaluation managed by MDRC, Barrow et al. (2012) find significant effects on credit accumulation for a subsample of Louisiana community college students enrolled at least half time. Early results from a series of similar randomized evaluations show small but statistically significant increases in cumulative earned credits by the first or second term (Cha and Patel 2010; Miller et al. 2011; Richburg-Hayes, Sommo, and Welbeck 2011). Evaluating a Canadian community college retention program, MacDonald et al. (2009) report significant increases in GPA and retention; this program paid $750 per semester for those with a GPA above 2.0, who maintained a full load *and* made use of academic services.

This paper reports results from an impact evaluation of a financial incentive demonstration program that builds on the lessons from earlier work. Overall academic performance in our study population was poor. Our merit aid therefore rewarded above-average performance for enrolled students. Specifically, the "Opportunity Knocks" (OK) experiment, piloted at a large Canadian commuter university, was designed to explore whether students who qualify for need aid can also be motivated by merit aid, and whether this improved performance would carry over into subsequent years. Rewarding higher grades in one year might generate better subsequent performance through habit formation or learning by doing, even after incentives disappear.

OK was offered to first- and second-year students who applied for financial aid. Those who signed up were randomly assigned to treatment and control groups. In con-

1. Randomized trials and quasi-experimental evaluations of financial incentives have been somewhat more encouraging for elementary and secondary students than for college students. Studies showing substantial positive effects on primary or secondary school students include Angrist et al. (2002), Henry and Rubinstein (2002), Kremer, Miguel, and Thornton (2009), Angrist and Lavy (2009), Dearden et al. (2009), Pallais (2009), and Dee (2011). Also in a primary or secondary context, Fryer (2012) reports large effects of aligned parent, teacher, and student incentives and Levitt et al. (2011) demonstrate some response to immediate rewards for test performance. Other recent experimental studies at this level have generated less reason for optimism. See, for example, Bettinger (2012), Rodriguez-Planas (2012), and Fryer (2011), who evaluate an array of award schemes for primary and middle school students in a variety of settings. For a general review of research on financial incentives, see Gneezy, Meier, and Rey-Biel (2011).

trast with earlier programs that rewarded students for achieving GPA thresholds, treated students earned $100 for *each* class in which they attained a grade of 70 or better and an additional $20 for each percentage point above 70 percent (roughly the average grade in the control group). A student with a full course load scoring 75 in every course qualified for $2,000 over the course of the school year (10 × ($100 + (5 × $20))). Treated students also had the opportunity to interact with randomly assigned peer advisors. These were upper-class students who had been trained to provide advice about study strategies, time management, and university bureaucracy.

OK was developed in view of the findings from our earlier randomized evaluation on a similar campus. The Student Achievement and Retention (STAR) project (Angrist, Lang, and Oreopoulos 2009) offered three interventions, the most successful of which combined financial incentives at widely spaced GPA thresholds with academic support services. OK provided an opportunity for replication and the chance to offer a more intense and perhaps even more successful treatment. By rewarding performance in each class and setting a low bar for the minimum payment, we hoped to make incentives stronger (92 percent of controls earned a grade of 70 percent or above in at least one class). This contrasts with STAR awards, which were paid out to only about 18 percent of eligible students. We opted for a partially linear payout scheme on theoretical grounds. (See, for example, Holmstrom and Milgrom 1987.)

OK awards were potentially more generous than those offered in STAR; high achievers could earn up to $700 *per class*.[2] The expected OK award among controls was $1,330, while the expected STAR award was only about $400. OK engendered more program engagement than STAR as well: Almost 90 percent of OK participants had some kind of interaction with peer advisors and/or the program website, in contrast with about 50 percent engagement in STAR.

OK had many novel and promising features: linear incentives at the class level, high reward levels, and high program engagement. It's therefore interesting, surprising, and somewhat disappointing that OK had only a modest impact on targeted outcomes. Treated second-year students earned about 13 percent more than expected based on the distribution of control-group grades, suggesting the program had an incentive effect. The strongest effects appear around the $100 award threshold, where completion of payment-qualifying courses increased, especially among students who appeared to understand the program well. OK also increased the number of second-year courses graded above 70 and grade points earned above 70, but these effects were not large enough to generate a significant increase in students' overall GPAs. OK generated no discernible impacts in the year after incentives were removed.

The following section describes the OK campus setting, program rules, and our random-assignment research design. Section III reports descriptive statistics and indicators of program engagement. Section IV discusses the experimental results while Section V reports on participants' impressions of the program as revealed in post-program surveys. The paper concludes in Section VI with a brief look at how our results fit in with other postsecondary incentive demonstrations. We also discuss possible explanations for differences between the findings reported here and those in our earlier study.

2. Tuition at this university is around $5,000 per year.

II. Background and Research Design

Motivated by the mixed results for college incentives to date, we developed an intervention meant to build on what we saw as the strongest features of the program studied by Angrist, Lang, and Oreopoulos (2009). The OK intervention combined incentives with academic support services; a combination of incentives and services appeared to be especially effective in the earlier STAR evaluation, which ran in a similar setting. The services delivered through STAR were more elaborate and expensive, however. STAR included the opportunity to participate in facilitated study groups as well as email-based peer mentoring, while OK services consisted of email-based peer mentoring only. We opted for email because the takeup rate for STAR's facilitated study groups was low. Also, because a number of STAR participants saw the awards as essentially out of reach, OK award rates were designed to be much higher. OK awards were also paid out more frequently, in this case, every term. Unlike STAR, the OK study population consisted only of students that had applied for financial aid prior to the start of the school year. This was partly in response to political constraints but it also seemed likely that aid recipients would be most responsive to the opportunity to earn additional awards.

Opportunity Knocks was piloted on an Ontario commuter campus affiliated with a large public university. The six-year completion rate on this campus is about 73 percent. There are about 2,500 students in an entering class. In late summer of 2008, we invited 1,056 first years and 1,073 second years to participate in OK. Eligible students are those who had requested financial aid, had an email address, had a high school GPA recorded in the university administrative information system, and who had enrolled for at least 1.5 credits for the upcoming fall term. Invitees who completed the intake survey and gave consent were eligible for random assignment. Of the 1,271 students who completed the survey and were eligible, 400 were treated. Treatment assignment was stratified by year (first and second) and sex, with 100 in each group. Within sex-year cells, assignment was stratified by high school GPA quartile, with 25 in each group. (The analysis below controls for strata.)

Previous studies have generally rewarded students for completing courses or reaching GPA thresholds. (See, for example, Angrist, Lang, and Oreopoulos 2009; Cha and Patel 2010.) In contrast, OK participants earned \$100 for each class in which they received at least a 70 percent grade, and an additional \$20 for each percentage point above 70.[3] For example, a student who earned a grade of 75 in each of five classes over one semester (five classes constitute a full load) would have received $5 \times (\$100 + (5 \times \$20)) = \$1{,}000$. We focused on grades near 70 because anything worse is typically seen as unsatisfactory and because awards for lower levels of achievement are likely to be prohibitively expensive (a GPA of at least C– is required for graduation; this translates to a percentage grade in the low 60s). Still, a grade of 70 is attainable for most students in at least one class, and the OK awards schedule provided incentives for above-average performance as well.

The services component of OK assigned treated students to (trained and paid) same-sex peer advisors. Peer advisors were enthusiastic upper-year students or recent graduates with good grades. Each peer advisor covered 50 participants. Advisors emailed advisees once every two to three weeks, whether or not the advisees responded. These

3. Payoffs were doubled and issued in the spring for year-long courses.

emails offered advice on upcoming academic events and workshops, and guidance relevant to key periods in the academic calendar, such as midterms and finals. Advisors also provided information about OK scholarships, including reminders of the scholarship calculation and payment schedules. Advisors frequently invited their clients to turn to them for help with any academic or personal issues that seemed relevant to academic success.

III. Descriptive Statistics and Program Response

The data for this study come primarily from the university records containing information on applicants, enrolled students, and course grades. We supplemented this with data from a baseline survey used to identify the population eligible for random assignment, as well as more descriptive focus-group style information collected from a few subjects after the experiment.

Table 1, which presents descriptive statistics, shows that OK participants were mostly college students of traditional age. Control group students had average grades around 82 percent in high school. Less than half of the control group spoke English as a first language, reflecting the relatively high proportion of immigrants on the OK host campus. About half of control group parents graduated from a postsecondary institution (44 percent of mothers and 53 percent of fathers), while nearly 80 percent of parents graduated from high school, a figure comparable to the Canadian average for college student parents. The OK scholarships were within reach for most participants: 92 percent of controls would have received an award under the OK scholarship formula. Table 1 also documents the fact that random assignment successfully balanced the background characteristics of those in the treatment and control groups (as evidenced by insignificant effects in the "Treatment Difference" columns). Although not documented in the table, student course selection and completion as measured by number of courses, difficulty, or subject area are also well balanced between treatment and control groups for the whole sample and within subgroups. (Random assignment occurred after students had preregistered for courses.)[4]

The OK intake survey, included in the packet describing the program to those eligible for random assignment, contained two questions meant to gauge subjects' understanding of program award rules. The first asked students to calculate the award amount for one class, and the second asked them to calculate the total award amount from five classes. Two-thirds of the students answered the second question correctly (documented in Table 1), and over 80 percent answered the first question correctly. Those who responded incorrectly to either question received a clarification by email. In the program analysis, we look at treatment effects for the entire sample and for those who answered the second assessment question correctly to see if those who understood the scholarship formula also had a stronger program response.

Student involvement with OK was high. This can be seen in Table 2, which shows that about 73 percent of treated students checked their scholarship earnings on the program website. Women were nine points more likely to check than men. Only 38 percent of treated participants sent an email to their assigned peer advisor in the fall,

4. Attrition was also balanced between treatment and control (about 5 percent of OK participants dropped out during the study), and treatment and control group dropouts have similar characteristics (results are available upon request).

Table 1
Descriptive Statistics and Covariate Balance by Gender

	Women				Men				All	
	First Years		Second Years		First Years		Second Years			
	Control Mean 1	Treatment Difference 2	Control Mean 3	Treatment Difference 4	Control Mean 5	Treatment Difference 6	Control Mean 7	Treatment Difference 8	Control Mean 9	Treatment Difference 10
Age	18.2	–0.105	19.2	0.011	18.4	0.014	19.2	0.069	18.7	–0.012
	[0.608]	(0.056)*	[0.514]	(0.056)	[0.815]	(0.104)	[0.460]	(0.070)	[0.757]	(0.036)
High school grade average	82.8	0.145	82.4	0.302	82.3	–0.344	82.1	–0.387	82.5	–0.024
	[6.56]	(0.238)	[6.19]	(0.217)	[6.44]	(0.310)	[6.73]	(0.338)	[6.44]	(0.134)
First language is English	0.404	0.057	0.426	–0.046	0.479	–0.060	0.333	0.097	0.416	0.009
	[0.491]	(0.056)	[0.495]	(0.057)	[0.501]	(0.065)	[0.474]	(0.069)	[0.493]	(0.031)
Mother a college graduate	0.395	0.065	0.477	–0.016	0.479	0.050	0.424	–0.034	0.439	0.020
	[0.490]	(0.056)	[0.500]	(0.058)	[0.501]	(0.065)	[0.497]	(0.070)	[0.496]	(0.031)
Father a college graduate	0.479	0.051	0.581	0.009	0.603	0.047	0.475	0.105	0.532	0.049
	[0.500]	(0.057)	[0.494]	(0.058)	[0.491]	(0.063)	[0.502]	(0.071)	[0.499]	(0.031)
Correctly answered harder	0.616	0.022	0.690	–0.010	0.719	–0.080	0.697	0.002	0.666	–0.014
question on scholarship formula	[0.487]	(0.053)	[0.464]	(0.054)	[0.451]	(0.061)	[0.462]	(0.065)	[0.472]	(0.029)
Controls who would have earned	0.883		0.968		0.908		0.978		0.923	
some scholarship money	[0.322]		[0.177]		[0.289]		[0.148]		[0.266]	
Hypothetical earnings for	1,240		1,390		1,430		1,400		1,330	
controls	[1,220]		[1,090]		[1,230]		[1,270]		[1,190]	
Observations	449		377		246		199		1,271	
F test for joint significance	1.11		0.453		0.858		1.43		0.515	
	{0.355}		{0.843}		{0.525}		{0.198}		{0.797}	

Notes: "Control Mean" columns report averages and standard deviations for variables in the left-most column, within the relevant gender-year subgroup. "Treatment Difference" columns report coefficients from regressions of each variable in the left-most column on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile). The last row presents within-column *F* tests of joint significance of all treatment differences. Control group standard deviations are in square brackets, robust standard errors are in parentheses, and *p* values for *F* tests are in curly braces. Some respondents did not answer the parents' education questions. They are coded as a separate category ("missing") and are not coded as high school or college graduates.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 2
Fraction of Treated Students Making Program-Related Contact by Gender and Year

	Women			Men			All		
Contact Type	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Emailed advisor (Fall)	0.450	0.390	0.420	0.410	0.270	0.340	0.430	0.330	0.380
Emailed advisor (Spring)	0.520	0.440	0.480	0.660	0.380	0.520	0.590	0.410	0.500
Emailed advisor (Fall or Spring)	0.790	0.700	0.745	0.750	0.560	0.655	0.770	0.630	0.700
Checked scholarship earnings online	0.760	0.780	0.770	0.650	0.710	0.680	0.705	0.745	0.725
Emailed the program website	0.270	0.320	0.295	0.250	0.300	0.275	0.260	0.310	0.285
Any contact	0.900	0.870	0.885	0.840	0.840	0.840	0.870	0.855	0.863
Observations	100	100	200	100	100	200	200	200	400

Notes: This table shows the proportion making the indicated form of program-related contact.

but this number increased to 50 percent in the spring. By year's end, 70 percent had emailed an advisor at least once over the course of the year. First-year students and women were more likely to contact advisors than were second-year students and men. At least 86 percent of treated students made some kind of program contact: They emailed a peer advisor, checked scholarship earnings, or emailed program staff.

Following a presentation of intention-to-treat effects, we discuss two-stage least squares (2SLS) estimates of treatment effects using a dummy indicating any program contact as the endogenous variable. The idea here is that subjects who made no program contact of any kind, and did not even check their scholarship earnings, are unlikely to have been affected by either OK awards or advisor services. In other words, we think of a dummy indicating any contact as a good surrogate for program treatment status. 2SLS estimates treating program contact as an endogenous variable should therefore capture the effect of treatment on the treated for the subpopulation of active program participants. (Because endogenous compliance is one-sided, the local average treatment effect is the treatment on the treated effect; see Imbens and Angrist 1994, for details.)

IV. Program Effects

A. Main Findings

A natural starting point for our analysis is a comparison of the amount earned by the experimental group with the earnings that students in the control group would have been entitled to had they been in the program. A large program effect should be reflected in larger-than-expected earnings, where expected earnings are measured using the grade distribution in the control sample.[5] Our estimates of earnings and other effects come from regressions like this one:

$$(1) \quad y_{ij} = \alpha_j + \beta T_i + \delta' X_i + \varepsilon_{ij},$$

where y_{ij} is the outcome for student i in stratum j, the α_j are strata effects, T_i is a treatment assignment indicator, and X_i is a vector of additional controls.[6] Causal effects of the OK program are captured by β. Because treatment is randomly assigned, covariates are unnecessary to reduce omitted variables bias in the estimated treatment effects. Models with covariates may, however, generate more precise estimates.

The OK program had no impact on earnings for first-year men and women, a result that can be seen in Columns 1, 4, and 7 of Table 3. On the other hand, there is some evidence of higher-than-expected earnings for second-year treated students, especially second-year men. The estimated effect on second-year men in the spring term, reported in Column 5, is a significant 170 dollars. Estimates over the course of the year are about 255 dollars for second-year men and 180 dollars for all second-years.[7] Both

5. Ashenfelter and Plant (1990) use a similar hypothetical payment outcome to measure the labor supply effects of exposure to a negative income tax.

6. Additional controls include parental education, an indicator for English mother tongue, and indicators for students who answered scholarship formula questions correctly.

7. Restricting the fall and spring samples to be the same as the full-year sample generates effects for the full year equal to the sum of the fall and spring effects. Estimated effects for the full year need not equal the sum (or average) of the two semester effects because the full-year sample differs slightly from the sample for either semester alone.

Table 3
Effects on (Hypothetical) Program Earnings

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel A. Fall									
Control mean	645	695	667	770	774	760	682	707	693
	[657]	[589]	[628]	[670]	[642]	[658]	[663]	[602]	[637]
Treatment effect	−18.8	99.7	39.9	33.9	49.2	11.9	−5.73	72.0	28.0
	(53.1)	(60.9)	(39.9)	(69.8)	(73.1)	(51.3)	(41.9)	(45.9)	(31.1)
N	444	374	818	246	195	441	690	569	1,259
Panel B. Spring									
Control mean	589	711	640	644	655	649	605	696	642
	[608]	[598]	[606]	[600]	[683]	[633]	[606]	[622]	[614]
Treatment effect	−57.6	24.7	−19.1	−20.0	170	35.5	−52.5	77.3	4.47
	(49.4)	(66.4)	(39.6)	(59.5)	(80.7)**	(49.4)	(37.6)	(51.0)	(30.8)
N	441	340	781	242	183	425	683	523	1,206

(*continued*)

Table 3 (*continued*)

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel C. Full Year									
Control	1,240	1,390	1,300	1,430	1,400	1,420	1,290	1,390	1,330
mean	[1,220]	[1,090]	[1,170]	[1,230]	[1,270]	[1,240]	[1,230]	[1,140]	[1,190]
Treatment	−80.2	165	33.0	7.01	255	54.8	−64.3	180	41.1
effect	(95.3)	(121)	(74.1)	(121)	(144)*	(95.2)	(74.3)	(91.3)**	(58.2)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Control Mean" rows list averages and standard deviations of program earnings, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions of program earnings on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Control earnings are hypothetical; treated earnings are actual. Full-year courses are double-weighted in the earnings calculation. The sample used for the full-year estimates includes students with grades in Fall and Spring. The Fall analysis omits full-year courses. If we restrict the Fall and Spring samples to be the same as the full-year sample, then the effects for the full-year are the sum of the Fall and Spring effects. (This is also true in later tables.) Robust standard errors are in parentheses; standard deviations are in square brackets.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

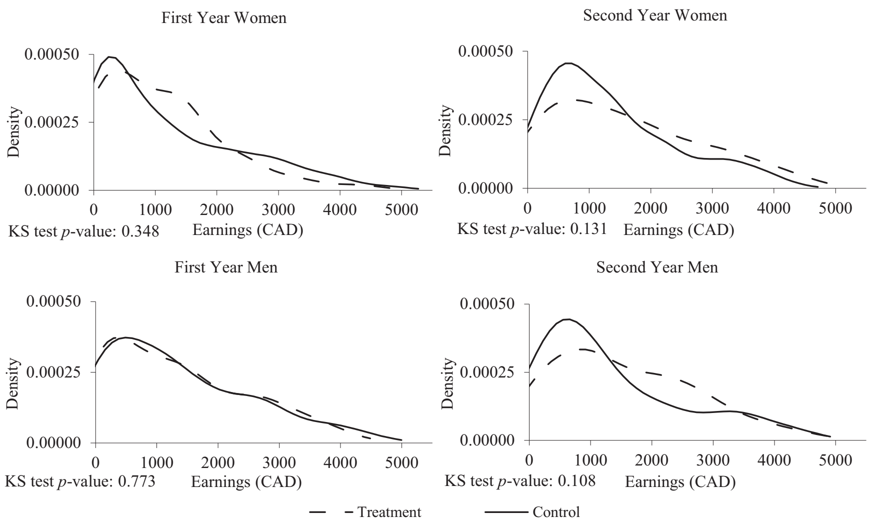

Figure 1
Densities of Full-Year Program Earnings

Note: The figure plots the smoothed kernel densities of OK program earnings for the 2008–9 school year. Control earnings are hypothetical; treated earnings are actual. Full-year courses are double-weighted in the earnings calculation. The sample used to make this figure includes students with grades in fall and spring.

of these estimates are significant at better than a 10 percent level and amount to 15–20 percent of a standard deviation of hypothetical control group earnings.

Our experimental design stratifies on sex, year of study, and high school GPA, mitigating concerns about mining for statistically significant findings in these subgroups. The analysis by sex and year is also of substantive interest. Still, it's worth noting that under the null hypothesis of no treatment effect for all four sex by class-year subgroups, the probability that at least one observed full-year effect is significant at the 8 percent level is $1 - 0.92^4 = 0.28$ (assuming no outcomes correlation across subgroups). As we show below, however, some results emerge more clearly when we limit the sample to students who understood the award formula well.

The question of whether the OK program caused more complex distributional shifts in hypothetical earnings is explored in Figure 1, which shows treatment and control earnings distributions in separate panels by sex and year. The only (marginally) significant distributional contrast in the figure is for second-year men (using a Kolmogorov-Smirnov test). On the other hand, the contrast by treatment status for second-year women looks similar to that for men. For both men and women, treatment seems to have shifted second-year earnings from below a level around 1,500 to more than 1,500 dollars. The shift emerges roughly one hundred dollars above mean earnings for controls.

The evidence for an effect on average grades (measured on a 0–100 scale) and GPA is weaker than that for earnings. The grades results appear in Table 4a and the GPA results appear in Table 4b. Average grades for second-year men increased by about 2.5 percentage points in the spring but this estimate is only marginally significant, and it's the only significant result in the table. The corresponding GPA effect amounts to about 0.27 GPA points, an estimate significant at the 5 percent level.[8] Power is not an issue with these comparisons. For the full sample, we are able to reject GPA and grade effects as small as 10 percent of the control standard deviation, meaning that our zeros are quite precise.

The earnings gains documented in Table 3 are necessarily explained by increases in the number of courses graded at least 70 and grade points over 70. Table 5 reports full-year program effects on each of these components of the scholarship award formula. Panel A shows effects on the number of courses in which a student earned a grade of at least 70. Treatment appears to have increased the number of over-70 grades awarded to second-year men by almost a full course. The number of over-70 courses increases by about half a course for all second-years. These estimates are reasonably precise. On the other hand, the estimated effects on grade points earned over 70 are not estimated very precisely. The only (marginally) significant point gain is for all second-years, an effect of 6.2 percentage points. It's also worth noting, however, that the magnitudes come out such that effects on total earnings are equally distributed between a threshold effect at 70 and awards for points over 70.

OK may have had a weaker effect on grades and GPA than on earnings because students substituted effort from classes with a grade above 70 to classes with a grade below 70. To test this claim and look for additional evidence of effects concentrated around the award threshold, we estimated treatment effects on indicators for grade>g, where g runs from 60 to 80 (reported in Figure 2; these plots also show the control grade distribution). This investigation uncovers no negative treatment effects on courses above the higher thresholds, suggesting that students generally did not substitute effort from higher- to lower-graded courses.[9]

We found no evidence of an increased likelihood of crossing any threshold for first-years. Treatment appears to have increased the likelihood that second-year women earned a grade of 72–74, a series of effects concentrated around the minimum award threshold. Effects concentrated around the threshold may be evidence of strategic grade-seeking behavior on the part of treated students. For example, students who expected a grade around 68 or 69 may have made a special effort (through negotiation or extra work) to clear 70. On the other hand, treatment appears to have boosted the grades of second-year men over a wide interval running from 60–75 percent. This pattern of effects weighs against a negotiation-based view of the incentive response, at least among men.

Although most students appeared to understand the OK program rules and award formula, a nontrivial minority did not. Those who misunderstood the formula linking grades and awards seem less likely to have been motivated by the awards. We therefore report estimates for a sample restricted to participants who correctly applied

8. GPA is not a linear transformation of average grades so we expect slight differences in results. Effects on GPA should be more similar to effects on earnings since GPA also jumps at 70 percent.

9. Similar analysis on courses graded above thresholds from 80 to 100 percent demonstrates little difference between treatment and control students.

Table 4a
Effects on Average Grades

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel A. Fall									
Control mean	68.1 [11.6]	71.0 [8.40]	69.4 [10.4]	70.7 [10.9]	72.4 [8.39]	71.4 [10.0]	68.9 [11.4]	71.4 [8.41]	70.0 [10.3]
Treatment effect	0.424 (0.945)	0.420 (0.947)	0.461 (0.662)	0.452 (1.18)	–0.520 (1.07)	–0.496 (0.827)	0.236 (0.740)	0.064 (0.694)	0.076 (0.515)
N	444	374	818	246	195	441	690	569	1,259
Panel B. Spring									
Control mean	67.4 [11.3]	71.2 [9.02]	68.9 [10.5]	68.8 [11.2]	70.0 [10.6]	69.3 [10.9]	67.8 [11.2]	70.8 [9.46]	69.0 [10.6]
Treatment effect	–0.814 (1.16)	–0.118 (1.13)	–0.471 (0.801)	–0.971 (1.56)	2.54 (1.41)*	0.106 (1.03)	–0.966 (0.901)	0.727 (0.901)	–0.225 (0.634)
N	441	340	781	242	183	425	683	523	1,206

(*continued*)

Table 4a (*continued*)

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel C. Full Year									
Control	67.9	71.1	69.2	69.9	71.5	70.5	68.4	71.2	69.6
mean	[10.7]	[7.77]	[9.69]	[10.3]	[8.59]	[9.70]	[10.6]	[7.99]	[9.70]
Treatment	–0.323	0.470	0.076	–0.233	1.17	–0.146	–0.458	0.614	–0.025
effect	(0.958)	(0.932)	(0.662)	(1.21)	(1.09)	(0.840)	(0.745)	(0.719)	(0.522)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Control Mean" rows list averages and standard deviations of average grades, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions of average grades on a treatment dummy, with sampling strata controls (year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Average grades are on a 100 point scale. Full-year courses are double-weighted in the average grade calculation. The sample used for the full-year estimates includes students with grades in Fall and Spring. The Fall analysis omits full-year courses. Robust standard errors are in parentheses; standard deviations are in square brackets.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 4b
Effects on GPA

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel A. Fall									
Control mean	2.39	2.64	2.50	2.61	2.75	2.66	2.46	2.67	2.55
	[0.982]	[0.765]	[0.900]	[0.920]	[0.743]	[0.856]	[0.968]	[0.760]	[0.890]
Treatment effect	0.021	0.046	0.038	0.046	–0.039	–0.034	0.014	0.015	0.009
	(0.079)	(0.081)	(0.056)	(0.103)	(0.098)	(0.073)	(0.063)	(0.061)	(0.044)
N	444	374	818	246	195	441	690	569	1,259
Panel B. Spring									
Control mean	2.34	2.64	2.47	2.47	2.54	2.50	2.38	2.61	2.48
	[0.916]	[0.783]	[0.875]	[0.935]	[0.880]	[0.912]	[0.922]	[0.810]	[0.885]
Treatment effect	–0.049	0.018	–0.016	–0.003	0.266	0.071	–0.037	0.102	0.022
	(0.081)	(0.090)	(0.059)	(0.106)	(0.119)**	(0.079)	(0.064)	(0.073)	(0.048)
N	441	340	781	242	183	425	683	523	1,206

(continued)

Table 4b (*continued*)

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel C. Full Year									
Control mean	2.37	2.64	2.49	2.55	2.67	2.59	2.42	2.65	2.52
	[0.895]	[0.689]	[0.825]	[0.870]	[0.739]	[0.822]	[0.890]	[0.702]	[0.825]
Treatment effect	–0.021	0.055	0.018	0.019	0.126	0.021	–0.019	0.075	0.019
	(0.073)	(0.079)	(0.053)	(0.096)	(0.097)	(0.070)	(0.058)	(0.061)	(0.042)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Control Mean" rows list averages and standard deviations of GPA, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions of GPA on a treatment dummy, with sampling strata controls (year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. GPA is on a four-point scale. The sample used for the full-year estimates includes students with grades in Fall and Spring. The Fall analysis omits full-year courses. Robust standard errors are in parentheses; standard deviations are in square brackets.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 5
Effects on Components of the OK Scholarship Formula

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel A. Number of Courses with Grade of At Least 70 Percent									
Control mean	4.58	5.22	4.85	5.18	5.01	5.11	4.75	5.16	4.92
	[3.35]	[2.84]	[3.16]	[3.17]	[2.96]	[3.08]	[3.30]	[2.87]	[3.14]
Treatment effect	–0.034	0.422	0.185	0.128	0.954	0.338	–0.010	0.572	0.239
	(0.260)	(0.335)	(0.205)	(0.356)	(0.405)**	(0.268)	(0.208)	(0.252)**	(0.161)
N	441	339	780	242	181	423	683	520	1,203
Panel B. Total Grade Percentage Points Over 70 Percent									
Control mean	38.9	43.3	40.8	45.5	45.0	45.3	40.9	43.8	42.1
	[46.2]	[42.1]	[44.5]	[47.4]	[50.4]	[48.5]	[46.6]	[44.4]	[45.7]
Treatment effect	–3.84	6.16	0.726	–0.290	7.98	1.05	–3.17	6.15	0.861
	(3.76)	(4.64)	(2.88)	(4.57)	(5.49)	(3.62)	(2.87)	(3.52)*	(2.25)
N	441	339	780	242	181	423	683	520	1,203

Notes: The dependent variable in Panel A is the total number of courses in which the student received a grade at 70 percent or higher over both semesters. In Panel B, the dependent variable is the sum of the percentage points by which the student's grades exceeded 70 percent. "Control Mean" rows list averages and standard deviations, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full-year courses are double-weighted in the calculation of both dependent variables. The sample used to make this table includes students with grades in Fall and Spring. Robust standard errors are in parentheses; standard deviations are in square brackets.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

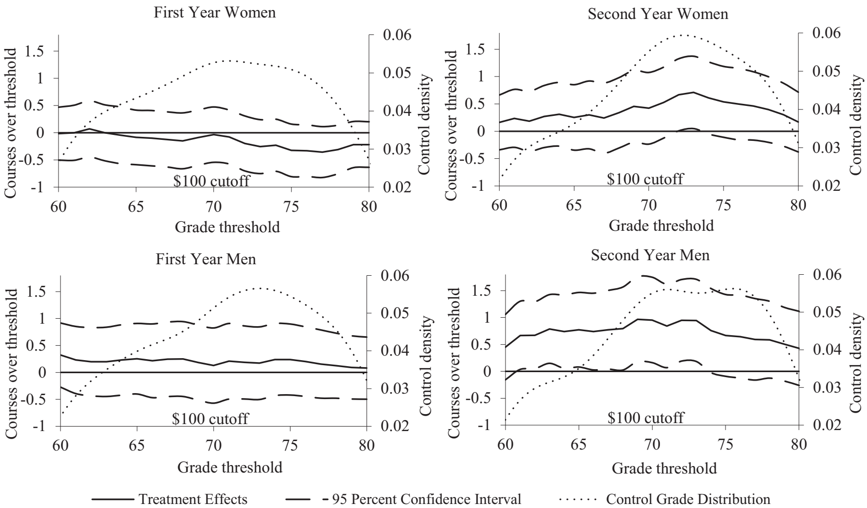

Figure 2
Full-Year Effects on Number of Courses with Grades Above Award Thresholds

Note: The figure shows treatment effects on the number of courses in which students earned a grade at or above a given threshold, where the thresholds are plotted on the x axis. Control densities are kernel density plots of grades at the course level using a normal kernel, taking only grades between 60 and 80 percent (inclusive). Treatment effects were estimated using the same models as for Table 3.

the OK earnings formula to an example in the baseline survey (information collected before random assignment). Two-thirds of the sample evaluated the example correctly.

Assuming that only those who understand the program change their behavior in response to OK incentives, average causal effects on those who understand program rules provide a measure of “theoretical effectiveness.” We’d expect to approach this bound over time, were schemes like OK a regular part of the college landscape. Estimates limited to the correct-responders sample are reported in Table 6.

Estimates for correct responders show larger program effects on earnings than the estimates computed using the full sample. Specifically, earnings gains are estimated to have been 370 for second-year men and 245 for all second-years, both significant at the 5 percent level. On the other hand, neither GPA nor grade effects are significantly different from zero in this sample. The apparent difference in findings for grades and earnings is explained by the last two rows of Table 6, which reports estimates for the components of the award formula in the restricted sample. These estimates show reasonably clear effects on the number of courses above 70 with weaker effects on points earned above. The shift in grades around the 70 percent threshold was apparently inadequate to boost overall GPA by a statistically significant amount.

Given the modest program effects observed during the treatment period, it seems unlikely that OK boosted achievement substantially in the longer-run. This conjecture is confirmed in Table 7, which reports full-sample treatment effects for fall 2009 (the semester after the program ended). The results in Table 7 show marginally significant

Table 6
Full-Year Effects (Students Who Calculated Awards Correctly)

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
(Hypothetical) program	–218	219	–9.32	102	370	160	–80.4	245	63.7
earnings	(130)*	(155)	(101)	(144)	(172)**	(111)	(97.2)	(114)**	(74.8)
Average grades	–1.23	0.999	–0.161	0.839	1.73	0.754	–0.351	1.03	0.219
	(1.10)	(1.12)	(0.779)	(1.51)	(1.31)	(1.00)	(0.913)	(0.879)	(0.634)
GPA	–0.107	0.112	–0.002	0.123	0.167	0.103	–0.008	0.117	0.044
	(0.088)	(0.095)	(0.064)	(0.118)	(0.117)	(0.083)	(0.072)	(0.074)	(0.052)
Number of courses with	–0.339	0.715	0.165	0.429	1.19	0.637	–0.008	0.813	0.353
grade of at least 70 percent	(0.333)	(0.410)*	(0.264)	(0.431)	(0.497)**	(0.323)**	(0.265)	(0.309)***	(0.203)*
Total grade percentage points	–9.21	7.38	–1.29	2.97	12.6	4.82	–3.98	8.19	1.42
over 70 percent	(5.25)*	(5.98)	(3.96)	(5.37)	(6.49)*	(4.19)	(3.81)	(4.37)*	(2.91)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Number of courses with grade of at least 70 percent" is the total number of courses in which the student received a grade at 70 percent or higher. "Total grade percentage points over 70 percent" is the sum of the percentage points by which the student's grades exceeded 70 percent. Each row reports coefficients from regressions of the indicated variable on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full-year courses are double-weighted in the calculation of the dependent variables. The sample used to make this table includes students with grades in Fall and Spring. Robust standard errors are in parentheses.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

positive effects on GPA for first-year women and average grades and GPA in the pooled sample of first-years (who are second-years in the posttreatment period), but these effects are small. The postprogram outcomes also offer a specification test for the analysis above since we would not expect to see threshold effects around 70 percent in the postprogram period. There is no evidence of a treatment effect on the number of fall 2009 courses graded at or above 70 percent.[10]

B. Subgroup Differences

Program effects differ by gender and year in school. First-years do not appear to have responded to the OK program while treated second-years—particularly second-year men—showed some improvement in grades, especially in courses graded over 70. Although we cannot be sure why results differ by sex and class, we hypothesize that first-years did not respond as strongly because many first-year students have not yet developed successful study techniques, raising their costs of grade improvement beyond OK's marginal returns.

The effect of marginal incentives might also depend on how well students can target their grades. For example, students who know they will fall just short of 70 might boost effort to clear this threshold. A student with inaccurate or imprecise grade knowledge may respond to the $100 payment even if his actual grades are well below or above 70. A possible explanation for the gender difference in our findings is a female advantage in grade targeting. Figure 2 shows localized positive treatment effects for second-year women around 72 to 73 percent, resulting in little effect on grades overall. Treated second-year men, however, increased courses graded above most thresholds from 60 to 75, contributing to stronger overall effects.

C. Additional Results

We might expect OK incentives to have been more powerful for financially constrained students. But treatment effects come out similar in subgroups defined by expected financial aid and whether students expressed concerns about funding. Effects are somewhat larger in the subsample of students whose parents had not been to college than among those with college-educated parents, but the gap by parents' schooling is not large or precisely estimated.

Effort substitution from easy to hard classes might have reduced the small treatment effects. OK participants had an incentive to switch to avoid hard classes. However, treatment effects do not vary by class difficulty, as measured by the average class grade among control students. (Results available upon request.) As noted above, OK also appears to have had little effect on course enrollment, difficulty, and completion.

The effects of program assignment reported in Tables 3 to 7 are diluted by noncompliance—that is, by the fact that some of those assigned to treatment did not really participate in the program because they were unaware of their assignment or uninterested in the program offerings. It's therefore worth estimating the effect of the

10. Roughly 100 program participants dropped out between the first and second years. Dropout rates were similar in the treatment and control groups.

Table 7
Effects in Fall 2009

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
(Hypothetical) program earnings	7.22	60.0	33.5	77.6	22.8	36.8	22.7	54.1	33.0
	(58.4)	(68.7)	(44.2)	(73.2)	(77.9)	(52.7)	(45.2)	(51.4)	(33.9)
Average grades	1.44	0.344	0.844	1.36	–2.16	–0.448	1.35	–0.618	0.299
	(0.917)	(1.17)	(0.736)	(1.49)	(1.46)	(1.06)	(0.803)*	(0.912)	(0.603)
GPA	0.148	0.019	0.082	0.083	–0.144	–0.037	0.119	–0.041	0.033
	(0.079)*	(0.096)	(0.062)	(0.127)	(0.122)	(0.088)	(0.068)*	(0.074)	(0.050)
Number of courses with grade of at least 70 percent	0.196	0.166	0.180	0.224	0.072	0.127	0.197	0.131	0.145
	(0.163)	(0.184)	(0.121)	(0.226)	(0.230)	(0.162)	(0.132)	(0.141)	(0.096)
Total grade percentage points over 70 percent	–0.620	2.17	0.776	2.76	0.782	1.21	0.152	2.05	0.921
	(2.32)	(2.69)	(1.75)	(2.74)	(3.02)	(1.99)	(1.75)	(2.02)	(1.32)
N	395	334	729	209	165	374	604	499	1,103

Notes: "Number of courses with grade of at least 70 percent" is the total number of courses in which the student received a grade at 70 percent or higher. "Total grade percentage points over 70 percent" is the sum of the percentage points by which the student's grades exceeded 70 percent. Each row reports coefficients from regressions of the indicated variable on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full-year courses are excluded from the calculation of all five dependent variables. "First-Year" and "Second-Year" continue to refer to the students' standing during the program period. Robust standard errors are in parentheses.

* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

scholarship and advisor treatment on program participants. The decision to engage with the program is not randomly assigned; this is a choice made by those offered the opportunity to participate. However, we can use the randomly assigned offer of OK treatment as an instrument for program takeup. By virtue of random assignment the OK offer is unrelated to characteristics of eligible students. The OK offer is also highly correlated with participation status: As shown in Table 2, 86 percent of those offered OK were engaged in some way—either through program/advisor contact or through checking scholarship earnings—while no one in the control group had access to OK awards or services. We assume that those with no program engagement were unaware of and therefore unaffected by the OK awards and services. The overall first-stage effect of OK offers on participation (awareness) is around 0.88, controlling for strata. (See Table 8.) Moreover, because no one in the control group participated, 2SLS estimates in this case capture the effect of treatment on the full sample of program participants, as described in Bloom (1984) and Imbens and Angrist (1994). Program participants are a self-selected group, but effects of OK on these students are of interest because they tell us how much achievement was boosted for those who were clearly aware of and responded to program opportunities in some measurable way.[11]

The first-stage effect of OK offers on participation rates is between 0.84 and 0.9 in the full sample and between 0.86 and 0.92 in the subsample that appears to have understood OK program rules. The first-stage estimates appear in the first row of each panel in Table 8, which also reports 2SLS estimates of the effect of participation on participants. Adjusting reduced-form offer effects (the estimates of program effects reported in Tables 3–6) for noncompliance necessarily leads to somewhat larger treatment effects, in this case larger by about 10–20 percent.

The most impressive effects in Table 8 are for the number of courses in which students earned a grade above 70. Here, effects on second-years in the full sample are on the order of two-thirds of a course, while the gains among those who understood the program well amount to almost a full course (an estimate of 0.91 with a standard error of 0.33, reported at the bottom of Column 8). The last column of Table 8 shows a marginally significant effect on the number of courses in which students earned at least 70 among all students who understood the program well (pooling men and women, and first- and second-years). The effect for all men is also significant at the 5 percent level in this sample, with a marginally significant impact on second-year women. A robust and substantial impact on hypothetical earnings and points above 70 also emerges from the 2SLS estimates in Panel B. At the same time, neither the earnings effects nor the increase in the number of courses graded above 70 translated into higher overall average grades among participants.

11. Some students may have been aware of the financial awards even though they failed to check their earnings or otherwise engage with the program. In this case, the reported first-stage effects on participation/awareness will be slightly too small, leading to inflated 2SLS estimates. Also, there is control noncompliance in the sense that control students have access to standard university support services. Therefore, the support services aspect of the OK treatment should be interpreted as a more engaging addition to a similar service rather than a new program implemented in a vacuum (Heckman et al. 2000).

Table 8
IV Estimates for Participants

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Panel A: Full Sample									
First stage (any contact)	0.901	0.891	0.897	0.844	0.874	0.858	0.876	0.882	0.878
	(0.029)***	(0.032)***	(0.022)***	(0.037)***	(0.035)***	(0.025)***	(0.023)***	(0.024)***	(0.017)***
Second stages:									
(Hypothetical)	–89.0	186	36.8	8.31	292	63.9	–73.4	204	46.8
Program earnings	(104)	(131)	(81.3)	(139)	(156)*	(108)	(83.6)	(101)**	(65.4)
Average grades	–0.359	0.527	0.084	–0.276	1.34	–0.171	–0.523	0.696	–0.029
	(1.05)	(1.02)	(0.727)	(1.38)	(1.18)	(0.956)	(0.840)	(0.795)	(0.587)
GPA	–0.023	0.062	0.020	0.023	0.144	0.024	–0.022	0.084	0.021
	(0.079)	(0.086)	(0.058)	(0.110)	(0.105)	(0.080)	(0.065)	(0.068)	(0.047)
Number of courses with grade of at least 70 percent	–0.037	0.473	0.206	0.152	1.09	0.394	–0.011	0.648	0.272
	(0.283)	(0.362)	(0.225)	(0.407)	(0.437)**	(0.304)	(0.234)	(0.277)**	(0.180)

(*continued*)

Table 8 (*continued*)

	Women			Men			All		
	First-Years 1	Second-Years 2	All 3	First-Years 4	Second-Years 5	All 6	First-Years 7	Second-Years 8	All 9
Total grade percentage points over 70 percent	–4.27	6.92	0.809	–0.344	9.14	1.22	–3.62	6.97	0.981
	(4.11)	(5.05)	(3.16)	(5.23)	(5.96)	(4.12)	(3.24)	(3.89)*	(2.53)
N	441	339	780	242	181	423	683	520	1,203
Panel B: Students Who Calculated Awards Correctly									
First stage (any contact)	0.922	0.907	0.915	0.863	0.900	0.875	0.896	0.895	0.895
	(0.033)***	(0.035)***	(0.024)***	(0.043)***	(0.037)***	(0.030)***	(0.027)***	(0.028)***	(0.019)***
Second stages:									
(Hypothetical) program earnings	–237	241	–10.2	119	411	183	–89.8	274	71.2
	(139)*	(164)	(108)	(158)	(178)**	(123)	(106)	(123)**	(82.0)
Average grades	–1.34	1.10	–0.176	0.972	1.92	0.862	–0.392	1.15	0.245
	(1.16)	(1.19)	(0.835)	(1.66)	(1.35)	(1.10)	(0.997)	(0.950)	(0.696)
GPA	–0.116	0.123	–0.002	0.143	0.186	0.117	–0.008	0.130	0.049
	(0.094)	(0.101)	(0.069)	(0.129)	(0.120)	(0.091)	(0.079)	(0.080)	(0.057)

Number of courses with grade of at least 70 percent	–0.368	0.788	0.181	0.497	1.32	0.729	–0.009	0.908	0.394
	(0.353)	(0.432)*	(0.282)	(0.475)	(0.511)**	(0.356)**	(0.289)	(0.332)***	(0.222)*
Total grade percentage points over 70 percent	–9.99	8.13	–1.41	3.45	14.0	5.51	–4.44	9.15	1.59
	(5.58)*	(6.34)	(4.25)	(5.91)	(6.71)**	(4.62)	(4.16)	(4.73)*	(3.19)
N	274	236	510	166	127	293	440	163	803

Notes: "First stage (any contact)" rows report coefficients from a regression of a dummy variable equal to one if the student made any program-related contact (see Table 2) on a treatment dummy. "Second stage" rows report coefficients from IV regressions, instrumenting for the program contact dummy with the treatment dummy. All regressions include sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full-year courses are double-weighted in the calculation of second-stage dependent variables. The sample used for this table includes students with grades in fall and spring. Standard errors are in parentheses.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

V. Student Impressions

The OK sign-up survey asked students to predict their average grades in two scenarios, one as an OK participant and one as a nonparticipant. To encourage a thoughtful response to this question, we offered those who answered the opportunity to win a $500 prize to be given to the student whose predictions came closest to the mark. About 60 percent predicted the same grade either way and the average predicted effect on grades was about 2.2 points. This is considerably larger than most of the effects reported in Tables 6 and 8. It also seems noteworthy that those who predicted a positive response do not appear to have responded more strongly than those who predicted no effect.

After the program ended, we asked students who predicted no effect in the intake survey why they had expected this. Of the 226 emails sent to treated participants predicting no effect, only 34 responded. Most of these respondents said they were planning to do as well as possible either way. For example, one said: “Before starting courses, I had already decided that I would do my best. And so, I felt a scholarship would be an added motivation, but fundamentally it came down to my own ability and commitment.” Two thought the award was too remote, commenting: “I predicted the program would have no effect because it provides a long-term reward for regular short-term behavior (daily intense studying).” Only three respondents said the incentives were too small. One said OK was “not too catchy and/or something worth dying for.” Another mentioned the 70 percent threshold: “I believe the cash reward for each course was not high enough per percentage point above 70 percent. If the cash reward was perhaps 30 or 40 dollars per percent point above 70 percent, I would’ve worked even harder.”

We also surveyed a random sample of 50 students from the treatment group at the end of the school year (May 13, 2009), offering $25 movie gift certificates to those who responded. Among the 30 respondents to this survey, 27 said the scholarships motivated them. Some thought the program was very effective. For example, one respondent commented: “Every time I began to lose interest in a particular course, I would remind myself that I just need to well . . . keep with it; the rewards will be tremendous. A scholarship is one such reward . . . and it sure is helpful, as it lifts a lot of the financial burdens I’m faced with when it comes to paying tuition & other fees.” Others saw the program was somewhat effective, as in this comment: “This scholarship did affect my motivation to study at some point . . .” Respondents often cited concerns about tuition and fees as motivating factors that boosted their interest in OK.

Half of the postprogram treated respondents felt the program led them to study more, though some felt their opportunity for more study time was limited. This comment was typical: “The program made me study more, but not much. I usually follow my schedule between work and school. So the amount of time I could have spent on study is somehow limited.” Others felt the program helped them focus on schoolwork: “As someone who gets sidetracked easily, I kept it in mind that staying focused would pay off in more than one way, and so yes, it did affect the amount of time I devoted to studying.” Another said, “I think what’s great about the program is that when you feel like you’re beginning to procrastinate, you think about the outcome of this program and want to get back to studying.” On the other hand, one second-year student reporting feeling somewhat demoralized by OK: “I did abnormally poor this year compared to my usual standards and it just so happened to coincide with Opportunity Knocks. The money reminder just kind of made me feel 'worse' about myself."

Among those who responded to the postprogram followup survey, almost all felt the program improved their academic performance. Some appreciated the opportunity to earn scholarships for good but not necessarily outstanding grades: "Personally, I don't find that [the university] offers as many scholarship opportunities as other [universities], so I think it was rewarding to know that my academic performance was acknowledged and rewarded." Some felt they increased performance out of financial concerns: "[E]specially now with the economic downfall, it is extremely difficult to muster up the finances to help pay for tuition without relying on OSAP [financial aid]. I kind of looked at Opportunity Knocks as my employer who gives me more money the better I performed in my studies." One student volunteered the view that the program would have a long-lasting effect on him/her: "The program had significantly improved my grades! And I cannot wait to see what I can accomplish next year."

Everyone we contacted afterward reported that they received peer advisor emails about once or twice a month. All but one of the respondents said the advisor emails were helpful. One noted, "I think the advisor made good decisions between sending us important reminders and information without being redundant. It was especially important to receive the emails about the scholarship money quickly after marks were sent in." Another said, "I find it very useful that someone was actually helping me through school." All but one respondent felt the program was worth continuing. Virtually everyone seemed grateful for having being selected for OK. One respondent closed with this endorsement: "The OK Program has been an essential part of my student experience, and in many ways crucial to my academic performance. I think that having a peer advisor as opposed to just the regular counselors working in the University is very important. With all the stress that universities cause their students—financially or otherwise, it's really nice to know there is a program like Opportunity Knocks to help students every step of the way." Overall, this feedback leaves us feeling that most treated students were aware of and engaged with OK, and that a large minority expected some benefit. Others who thought the program would have little effect seem to feel this way because they were already anxious to succeed and willing to devote time to their studies.

VI. Summary and Conclusions

The OK program was popular with participants: Sign-up rates and program engagement were high, and in focus group interviews many program participants were enthusiastic about their experiences. This enthusiasm probably reflects the high award rates for OK. It's therefore disappointing that, despite the introduction of substantial awards at almost every relevant level of achievement, overall program effects on achievement were modest. On the plus side, treated second-year students earned more in OK scholarship money than we would have expected based on the control-group grade distribution, increased the number of courses in which they earned a grade of 70, and gained a few grade points above 70. This localized response did not translate into a substantial boost in overall achievement, though it was noticeably stronger in the subsample of students who appear to have understood the OK award scheme well.

The past decade has seen a growing number of randomized evaluations of pay-for-performance schemes for students at various levels and quasi-experimental studies looking at effects of state-based merit aid programs. Table 9 summarizes studies using randomized designs to look at financial incentives in college and Table 10 lists results from quasi-experimental studies of state-based merit scholarships.[12] A number of randomized evaluations show effects on credits earned in response to incentives for course completion and grade thresholds. (See, for example, Cha and Patel 2010; Barrow et al. 2012.) These results, along with the findings in Angrist, Lang, and Oreopoulos (2009) and those reported here suggest that students react to threshold targets more strongly than to marginal incentives beyond the initial target. Our linear incentive scheme was characterized by a fairly forgiving initial target, a fact that may have induced a stronger threshold response. The OK program's novel linear incentive of $20 per percentage point provides a lower bound (in this context at least) for the marginal incentive needed to induce substantially higher student effort, especially for first years.

We were also surprised when the OK demonstration failed to replicate the strong positive results for women seen in the STAR experiment. Women may have shown a weaker, localized response to OK because they successfully targeted the initial award cutoff. Men do not appear to have responded as strategically as women to the OK incentive scheme, yet they did not respond to STAR's incentives to any degree. Perhaps the STAR GPA awards were simply too uncertain or too small to motivate men.

Incentives seem to be more effective when combined with academic support services. On balance, however, the picture that emerges from Table 9 and from this study is one of mostly modest effects. In particular, overall GPA seems largely unaffected except in some subgroups, and Angrist, Lang, and Oreopoulos (2009) is the only randomized evaluation to date showing college achievement effects persisting into the posttreatment period. Table 10 describes similarly discouraging results from studies of state-based merit aid programs. A few studies report positive effects, most notably Scott-Clayton's (2011) evaluation of West Virginia PROMISE. However, other positive results appear weaker in light of updated empirical work (Sjoquist and Winters 2012a, 2012b) and a better understanding of selection effects (Cohodes and Goodman forthcoming).

The muted effectiveness of merit scholarships may partly reflect the trouble struggling students have developing effective study strategies. For example, Israeli high school students have easy access to test-focused study sessions in public school, a fact that may explain some of the stronger Angrist and Lavy (2009) results on achievement awards for high school girls. Indeed, second-year students may have responded more strongly in our study precisely because they have a better sense of how to improve their grades. Fryer (2011) similarly argues that incentives for learning (in his case, reading books) look more promising than pay for performance on achievement tests. These intriguing results come from elementary and secondary school settings. Investigation of the merits of as-yet-untried recipes combining learning incentives

12. The studies listed in Table 9 use random assignment to evaluate financial incentives for college students. This list is the result of a citation search (that is, citing studies we were previously aware of), a keyword search (for "experiment, incentives, college") using Google Scholar, and helpful suggestions from anonymous referees. Table 10 was constructed similarly based on studies using difference in differences, regression discontinuity, event study designs to test impacts of state-based merit aid programs on college performance and completion.

Table 9
Randomized Evaluations of College Achievement Awards

	Study	Sample	Treatment	Outcome	Effects		
					All	Men	Women
	1	2	3	4	5	6	7
1.	Angrist, Lang, and Oreopoulos (2009) [The Student Achievement and Retention Project]	First year students at Canadian commuter university in 2005– 2006, except for top HS grade quartile	$1,000 for C+ to B– first year performance, $5,000 for B+ to A performance (varies by HS grade)	First year GPA	0.010 (0.064) [1.805]	–0.110 (0.103) [1.908]	0.086 (0.084) [1.728]
				First year credits earned	–0.012 (0.064) [2.363]	–0.157 (0.106) [2.453]	0.084 (0.082) [2.298]
				Second year GPA	–0.018 (0.066) [2.040]	–0.081 (0.108) [2.084]	0.030 (0.085) [2.008]
				Second year credits earned	0.027 (0.108) [2.492]	0.155 (0.180) [2.468]	–0.024 (0.137) [2.509]
			Incentives and support services	First year GPA	0.210 (0.092)** [1.805]	0.084 (0.162) [1.908]	0.267 (0.117)** [1.728]
				First year credits earned	0.092 (0.087) [2.363]	–0.196 (0.015) [2.453]	0.269 (0.108)** [2.298]
				Second year GPA	0.072 (0.091) [2.040]	–0.170 (0.161) [2.084]	0.276 (0.106)*** [2.008]

(*continued*)

Table 9 (*continued*)

	Study 1	Sample 2	Treatment 3	Outcome 4	Effects: All 5	Effects: Men 6	Effects: Women 7
				Second year credits earned	0.072 (0.130) [2.492]	–0.240 (0.206) [2.468]	0.280 (0.172) [2.509]
2.	Angrist, Oreopoulos, and Williams (2013) [Opportunity Knocks]	First-year students on financial aid at Canadian commuter university in 2008–2009	Over two semesters and for each semester-long course, $100 for attaining at least 70 percent and $20 for each percentage point higher than this (full course load = 10 semester courses)	First year GPA	–0.019 (0.058) [2.42]	0.019 (0.096) [2.55]	–0.021 (0.073) [2.37]
				GPA, fall term of year after program	0.119 (0.068)* [2.60]	0.083 (0.127) [2.58]	0.148 (0.079)* [2.61]
		Second-year students on financial aid at Canadian commuter University in 2008–2009		First year GPA	0.075 (0.061) [2.65]	0.126 (0.097) [2.67]	0.055 (0.079) [2.64]
				GPA, fall term of year after program	–0.041 (0.074) [2.83]	–0.144 (0.122) [2.79]	0.019 (0.096) [2.85]
3.	Barrow et al. (2012) [Opening Doors Louisiana]	Low-income parents beginning community college in Louisiana between 2004 and 2005	For each of two semesters, $250 for at least half-time enrollment, $250 for C average or better at end of midterms, and $500 for maintaining a C average, plus optional enhanced college counseling	First semester credits earned	1.222 (0.285)*** [4.609]		
				Second semester credits earned	1.126 (0.265)*** [2.77]	sample is mostly female	
				Credits earned, year after program	0.343 (0.456)		

4.	Cha and Patel (2010) [Ohio Performance-Based Scholarship Demonstration]	Low-income Ohio college students in 2008 with children and eligible for TANF	$1,800 for earning a grade of C or better in 12 or more credits, or $900 for a C or better in 6 to 11 credits, with payments at end of each semester	First year credits attempted	0.5 (0.4) [19.5]	sample is mostly female	
				First year credits earned	2.0 (0.5)*** [13.4]		
5.	De Paola, Scoppa, and Nistico (2012)	First-year business students at the University of Calabria in 2008–2009	$1,000 for students with the 30 highest cumulative scores on all exams	Cumulative exam score	6.023 (3.059)**	5.390 (4.615)	5.841 (4.061)
				Credits earned	2.335 (1.197)**	1.759 (1.854)	2.490 (1.518)*
			$350 for students with the 30 highest cumulative scores on all exams	Cumulative exam score	5.350 (3.164)*	2.354 (4.877)	6.157 (4.207)
				Credits earned	2.194 (1.266)*	0.714 (1.970)	2.766 (1.655)*
6.	Leuven, Oosterbeek, and van der Klaauw (2010)	First-year economics and business students at the University of Amsterdam in 2001–2002	$600 for completion of all first year requirements by start of new academic year	Met first year requirements	0.046 (0.065) [0.195]	not reported	
				Total “credit points” in first three years	–1.2 (9.8) [84.3]		
			$200 for completion of all first year requirements by start of new academic year	Met first year requirements	0.007 (0.062) [0.195]		
				Total “credit points” in first three years	–2.5 (9.6) [84.3]		

(*continued*)

Table 9 (*continued*)

	Study 1	Sample 2	Treatment 3	Outcome 4	Effects All 5	Effects Men 6	Effects Women 7
7.	Leuven et al. (2011)	First-year economics and business students at the University of Amsterdam in 2004–2005 and 2005–2006	$1,250 for the student with the top microeconomics exam score	Microeconomics exam score	0.974 (0.877) [18.7]		
			$3,750 for the student with the top microeconomics exam score	Microeconomics exam score	1.184 (0.617)* [18.9]	not reported	
			$6,250 for the student with the top microeconomics exam score	Microeconomics exam score	–0.629 (0.644) [21.2]		
8.	MacDonald et al. (2009) [Foundations for Success]	At-risk students beginning community college in Ontario, Canada, between 2007 and 2008	$750 each of three semesters for 1) obtaining 2.0 GPA or higher, 2) eligible to continue in a full program the following semester, and 3) completing at least 12 hours of tutorial, case management, or career workshops	First semester GPA during program (missing imputed)	0.07 $p > 0.1$ [2.11]		0.12 $p > 0.1$ [2.20]
				Second semester GPA during program (missing imputed)	0.12 $p < 0.05$** [1.88]	not reported	0.14 $p < 0.05$** [2.04]
				Third semester GPA during program (missing imputed)	0.01 $p > 0.1$ [2.09]		0.12 $p < 0.05$** [2.16]
				Fourth semester GPA (after program)	0.07 $p > 0.1$ [2.18]		0.16 $p < 0.05$** [2.33]

9.	Miller et al. (2011) [New Mexico Performance-Based Scholarship Demonstration]	Low-income students starting at the University of New Mexico in fall, 2008, and fall, 2009	$1,000 each of four semesters for 1) obtaining 2.0 GPA or higher, 2) enrolling full time, and 3) completing two extra advisor meetings per semester	First semester credits earned	0.0 (0.2) [12.8]	not reported
				Second semester credits earned	0.6 (0.3)* [11.1]	
10.	Richburg-Hayes Sommo, and Welbeck (2011) [New York Performance-Based Scholarship Demonstration]	New York City community college students aged 22–35 who required remediation fall, 2008, through fall, 2009	Up to $1,300 each of two or three semesters, paid in installments for achieving 1) registration, 2) continued mid-semester enrollment, and 3) 2.0 GPA in at least six credits	First semester credits earned	0.6 (0.3)* [8.1]	not reported
				Second semester credits attempted	0.6 (0.4) [9.3]	

Notes: The table reports treatment effects for grades, credits earned, and measures of persistence. Standard errors are shown in parentheses. Control means are shown in square brackets.
* significant at 10 percent level. ** significant at 5 percent level. *** significant at 1 percent level

Table 10
Quasi-Experimental Evaluations of Merit-Based College Scholarships

	Study 1	Treatment 2	Methodology 3	Outcome 4	Effects 5
1.	Castleman (2014) [Bright Futures Scholarship, Florida Medallion Scholars (FMS) and Florida Academic Scholars (FAS)]	FMS: 75 percent of public college tuition and fees for students with a 3.0 high school GPA and at least 20 on the ACT or 970 on the SAT	Difference in differences, non-eligible students as controls, HS graduates in 2000	FL public college credits, four years	–0.87 (1.90)
				FL public college BA in four years	0.00 (0.02)
		FAS: 100 percent of public college tuition and fees for students with a 3.5 high school GPA and at least 28 on the ACT or 1270 on the SAT	Difference in differences, non-FMS students as controls, HS graduates in 2000	FL public college credits, four years	9.05 (1.75)***
				FL public college BA in four years	0.07 (0.02)***
2.	Cohodes and Goodman (Forthcoming) [John and Abigail Adams Scholarship Program (MA)]	MA public school tuition waived (excluding substantial fees) for students who score in the top 25th percentile of their school district and attain minimum absolute benchmarks on the statewide tenth grade test; must maintain 3.0 GPA in college	Regression discontinuity on tenth grade test score	Enrolled on time at a four-year college	0.009 (0.008) [0.716]
				Graduated in four years from a four-year college	–0.017 (0.010)* [0.433]
3.	Cornwell, Lee, Mustard (2005) [Georgia HOPE]	Full tuition/fees at GA public colleges for students with a 3.0 high school GPA; must maintain 3.0 GPA in college	Differences in differences, non-GA-resident students as controls	Enrolled in full freshman course load at University of Georgia	–0.042 (0.016)*** [0.812]
				Completed full freshman course load at University of Georgia	–0.060 (0.019)*** [0.588]

4.	Dynarski (2008) [Georgia HOPE and Arkansas merit aid program]	$1,000 at inception (now $2,500) for tuition/fees at AR colleges for students with at least 19 on the ACT and a 2.5 core high school GPA; full tuition/fees at GA public colleges for students with a 3.0 high school GPA; for AR and GA, must maintain 3.0 GPA in college	Difference in differences, other state populations as controls	Fraction of age 22–34 population with a college degree	0.0298 (0.004)*** [0.337]
5.	Scott-Clayton (2011) [West Virginia PROMISE]	Full tuition/fees at WV public colleges for students with a 3.0 overall and core high school GPA and at least 21 on the ACT or 1000 on the SAT	Regression discontinuity on ACT score	Four-year GPA, WV public college students Earned BA in four years	0.099 (0.045)** 0.094 (0.022)**
			Event study, program introduction (small sample T-distribution critical values)	Four-year GPA, WV public college students Earned BA in four years	0.039 (0.018) 0.067 (0.005)***
6.	Sjoquist and Winters (2012a) [Georgia HOPE and Arkansas merit aid program]	$1,000 at inception (now $2,500) for tuition/fees at AR colleges for students with at least 19 on the ACT and a 2.5 core high school GPA; full tuition/fees at GA public colleges for students with a 3.0 high school GPA; for AR and GA, must maintain 3.0 GPA in College	Difference in differences, other state populations as controls; increased sample and updated clustering compared with Dynarski (2008)	Fraction of age 22–34 population with a college degree	0.0091 p=0.216 [0.3567]
7.	Sjoquist and Winters (2012b)	25 state merit aid programs with requirements on high school GPA, ACT/SAT scores, and college credit enrollment and GPA	Difference in differences, non-merit state populations as controls	Fraction of age 24–30 population with a college degree	–0.0008 (0.0028) [0.388]
		Nine strongest state merit aid programs with requirements on high school GPA, ACT/SAT scores, and college credit enrollment and GPA		Fraction of age 24–30 population with a college degree	0.0011 (0.0037) [0.388]

Notes: The table reports main baseline sample outcomes for grades and measures of persistence. Standard errors are shown in parentheses. Control means are shown in square brackets.
* significant at 10 percent level. ** significant at 5 percent level. *** significant at 1 percent level

with academic support schemes seems a worthy priority for future research on college achievement.

Our study also indicates that program awareness and understanding are important features of college incentive design. The positive effects of OK, though muted, are concentrated among students who understood the awards formula well. And adjusting for the fact that 14 percent of students assigned to treatment did not engage with the program in any way generates larger estimates for those who were aware of the program. Program effects may therefore increase as program awareness and understanding increase.

References

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. “Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment.” *American Economic Review* 92(5):1535–58.

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

Angrist, Joshua D., and Victor Lavy. 2009. “The Effect of High School Matriculation Awards: Evidence from a Randomized Trial.” *American Economic Review* 99(4):1384–414.

Arum, Richard, and Josipa Roksa. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago Press.

Ashenfelter, Orley, and Mark W. Plant. 1990. “Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs.” *Journal of Labor Economics* 8(1, Part 2: Essays in Honor of Albert Rees):S396–415.

Babcock, Philip, and Mindy Marks. 2011. “The Failing Time Cost of College: Evidence from Half a Century of Time Use Data.” *Review of Economics and Statistics* 93(2):468–78.

Barrow, Lisa, Lashawn Richburg-Hayes, Cecilia Elena Rouse, and Thomas Brock. 2012. “Paying for Performance: The Educational Impacts of a Community College Scholarship Program for Low-Income Adults. Working Paper, Federal Reserve Bank of Chicago.

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

Bettinger, Eric, and Rachel Baker. 2011. “The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring.” NBER Working Paper No. 16881. Cambridge: National Bureau of Economic Research.

Bloom, Howard S. 1984. “Accounting for No-Shows in Experimental Evaluation Designs.” *Evaluation Review* 8(2):225–46.

Bound, John, Michael F. Lovenheim, and Sarah Turner. 2010. “Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources.” *American Economics Journal: Applied Economics* 2(3):129–57.

Castleman, Benjamin L. 2014. “The Impact of Partial vs. Full Merit Scholarships on College Entry and Success: Evidence from the Florida Bright Futures Scholarship Program.” EdPolicyWorks Working Paper Series No. 17.

Cha, Paulette, and Reshma Patel. 2010. “Rewarding Progress, Reducing Debt: Early Results from Ohio’s Performance-Based Scholarship Demonstration for Low-Income Parents.” MDRC Report, October. New York and Oakland: MDRC.

Cohodes, Sarah, and Joshua Goodman. Forthcoming. “Merit Aid, College Quality, and College Completion: Massachusetts’ Adams Scholarship as an In-Kind Subsidy.” *American Economics Journal: Applied Economics*, Forthcoming.

Cornwell, Christopher, Kyung Hee Lee, and David B. Mustard. 2005. "Student Responses to Merit Scholarship Retention Rules." *Journal of Human Resources* 40(4):895–917.
Cornwell, Christopher, and David B. Mustard. 2007. "Merit-Based College Scholarships and Car Sales." *Education Finance and Policy* 2(2):133–51.
Cornwell, Christopher, David B. Mustard, and Deepa J. Sridhar. 2006. "The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program." *Journal of Labor Economics* 24(4):761–86.
De Paola, Maria, Vincenzo Scoppa, and Rosanna Nistico. 2012. "Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment." *Journal of Human Capital* 6(1):56–85.
Dearden, Lorraine, Carl Emmerson, Christine Frayne, and Costas Meghir. 2009. "Conditional Cash Transfers and School Dropout Rates." *Journal of Human Resources* 44(4):827–57.
Dee, Thomas. 2011. "Conditional Cash Penalties in Education: Evidence from the Learnfare Experiment." *Economics of Education Review* 30(5):924–37.
Dynarski, Susan. 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):1755–98.
———. 2012. "Aligning Student, Parent and Teacher Incentives: Evidence from Houston Public Schools." Harvard University, Working Paper.
Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Retorre. 2012. "College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities." *Review of Economics and Statistics* 94(3):699–711.
Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25(4):191–210.
Harris, Douglas N., and Sara Goldrick-Rab. 2010. "The (Un)Productivity of American Higher Education: From 'Cost Disease' to Cost-Effectiveness." University of Wisconsin, Working Paper.
Henry, Gary T., and Ross Rubinstein. 2002. "Paying for Grades: Impact of Merit-Based Financial Aid on Education Quality." *Journal of Policy Analysis and Management* 21(1): 93–109.
Heckman, James, Neil Hohmann, Jeffrey Smith, and Michael Khoo. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *Quarterly Journal of Economics* 115(2):651–94.
Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55(2):303–28.
Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2):467–75.
Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *The Review of Economics and Statistics* 91(3):437–56.
Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. 2010. "The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment." *Journal of the European Economic Association* 8(6):1243–65.
Leuven, Edwin, Hessel Oosterbeek, Joep Sonnemans, and Bas van der Klaauw. 2011. "Incentives Versus Sorting in Tournaments: Evidence from a Field Experiment." *Journal of Labor Economics* 29(3):637–58.
Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff. 2011. "The Impact of Short-Term Incentives on Student Performance." University of Chicago, Working Paper.
MacDonald, Heather, Robert Malatest, Rob Assels, Rana Baroud, Lili Gong, Larry Bernstein, Cristofer Price, and John Greenwood. 2009. "Final Impacts Report: Foundations for Success Project." Report to the Canada Millennium Scholarship Foundation. Ottawa: Canada Millennium Scholarship Foundation.
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 Mexico." MDRC Report, August. New York and Oakland: MDRC.
Pallais, Amanda. 2009. "Taking a Chance on College: Is the Tennessee Education Lottery-Scholarship Program a Winner?" *The Journal of Human Resources* 44(1):199–222.
Richburg-Hayes, Lashawn, Colleen Sommo, Rashida Welbeck. 2011. "Promoting Full-Time Attendance among Adults in Community College: Early Impacts from the Performance-Based Scholarship Demonstration in New York." MDRC Report, May. New York and Oakland: MDRC.
Rodriguez-Planas, Nuria. 2012. "Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States." *American Economic Journal: Applied Economics* 4(4): 121–39.
Scrivener, Susan, Colleen Sommo, and Herbert Collado. 2009. "Getting Back on Track: Effects of a Community College Program for Probationary Students." MDRC Report, April.
Scrivener, Susan, and Michael J. Weiss. 2009. "More Guidance, Better Results? Three-Year Effects of an Enhanced Student Services Program at Two Community Colleges." MDRC report, August. New York and Oakland: MDRC.
Scott-Clayton, Judith. 2011. "On Money and Motivation: A Quasi-Experimental Analysis of Financial Incentives for College Achievement." *Journal of Human Resources* 46(3):614–46.
Shaienks, Danielle, and Tomasz Gluszynski. 2007. "Participation in Postsecondary Education: Graduates, Continuers, and Dropouts, Results from YITS Cycle 4." Research Paper, Culture, Tourism and the Centre for Education Statistics. Ottawa: Statistics Canada.
Shapiro, Doug, Afet Dundar, Jin Chen, Mary Ziskin, Eunkyong Park, Vasti Torres, Yi-Chen Chiang. 2012. "Completing College: A National View of Student Attainment Rates." Signature Report 4, National Student Clearinghouse Research Center. Herndon: National Student Clearinghouse.
Sjoquist, David L., and John V. Winters. 2012a. "Building the Stock of College-Educated Labor Revisited." *Journal of Human Resources* 47(1):270–85.
———. 2012b. "State Merit-Based Financial Aid Progams and College Attainment," IZA Discussion Paper No. 6801. Bonn, Germany: Institute for the Study of Labor.
Turner, Sarah E. 2004. "Going to College and Finishing College: Explaining Different Educational Outcomes." In *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, ed. Catherine M. Hoxby, 13–56. Chicago: University of Chicago Press.