American Economic Journal: Applied Economics 2009, 1:1, 136–163
http://www.aeaweb.org/articles.php?doi=10.1257/app.1.1.136

Incentives and Services for College Achievement: Evidence from a Randomized Trial†

By Joshua Angrist, Daniel Lang, and Philip Oreopoulos*

This paper reports on an experimental evaluation of strategies designed to improve academic performance among college freshmen. One treatment group was offered academic support services. Another was offered financial incentives for good grades. A third group combined both interventions. Service use was highest for women and for subjects in the combined group. The combined treatment also raised the grades and improved the academic standing of women. These differentials persisted through the end of second year, though incentives were given in the first year only. This suggests study skills among some treated women increased. In contrast, the program had no effect on men. (*JEL* I21, I28)

Recent years have seen growing interest in interventions designed to increase college attendance and completion, especially for low-income students. Major efforts to increase enrollment include need-based and merit-based aid, tax deferral programs, tuition subsidies, part-time employment assistance, and improvements to infrastructure. The resulting expenditures are justified in part by empirical evidence which suggests that there are substantial economic returns to a college education and to degree completion in particular (see, e.g., Thomas J. Kane and Cecilia Elena Rouse 1995).

In addition to the obvious necessity of starting college, an important part of the post-secondary education experience is academic performance. Many students struggle and take much longer to finish than the nominal completion time. First-year students are especially likely to run into trouble. Nearly one-third of first-year

* Angrist: Department of Economics, Massachusetts Institute of Technology, E52-353, 50 Memorial Drive, Cambridge, MA 02142-1347 (e-mail: angrist@mit.edu); Lang: Ontario Institute for Studies in Education, The University of Toronto, 252 Bloor Street West, Toronto, ON M5S 1V6 (e-mail: dlang@oise.utoronto.ca); Oreopoulos: Department of Economics, University of British Columbia, 997-1873 East Hall, Vancouver, BC V6T 1Z1, and University of Toronto National Bureau of Economic Research and Canadian Institute for Advanced Research (e-mail: oreo@economics.utoronto.ca). This project was sponsored by the Canada Millennium Scholarship Foundation. Special thanks go to the staff and administrators who assisted with the project and to Cynthia Kinnan and Simone Schaner for exceptional research assistance. We benefited from helpful comments by seminar participants at the National Bureau of Economic Research Summer Institute, the Stanford Institute Theoretical Conference (SITE), the Canadian Institute for Advanced Research, Statistics Canada, Stanford University, Harvard University, Rice University, Texas A&M, Massachusetts Institute Technology, the London School of Economics, the Paris School of Economics, National Taiwan University, the University of Washington, University of Toronto, University of Houston, University of Maryland at College Park, St. Gallen, and the University of California, Santa Cruz, University of California, Santa Barbara, and University of California, Berkeley.

† To comment on this article in the online discussion forum visit the articles page at: http://www.aeaweb.org/articles.php?doi=10.1257/app.1.1.136

college students in the United States participate in remedial courses in reading, writing, or mathematics (National Center for Education Statistics 2003). About one in five students who begin a four year college program leave within a year, either voluntarily or because of unsatisfactory achievement. About two in five leave within six years without a degree (Consortium for Student Retention Data Exchange 2004).

One reason for poor student performance is lack of preparation. In particular, many students have poor study skills. Motivated by the view that the return to these skills is high, the traditional response to achievement problems has been an array of academic service strategies (Betsey O. Barefoot 2004). For example, most North American institutions offer note taking, time management, and goal setting workshops, as well as academic advising and remedial instruction. Sometimes academic support services are combined with psychological support services (Vincent Tinto 1993, John I. Goodlad 2004).

Like academic support services, merit scholarships have a long history in the post-secondary context, but traditional programs, such as the US National Merit awards and Canadian Excellence Awards, have focused on a small number of very high achievers.[1] A recent development in the scholarship field is an attempt to use financial awards and incentives to motivate good but not spectacular students. Examples such as Georgia's Helping Outstanding Pupils Educationally (HOPE) program include state tuition waivers for students who maintain a B average. As Susan Dynarski (2005) notes, programs of this sort are relevant for many students. Nearly 60 percent of high school graduates in Georgia qualify for a HOPE scholarship (if they go to college). In addition to providing more financial resources for college, a second goal of the HOPE program is to promote academic achievement (Jason Seligman et al. 2004). The promise of a scholarship may increase the time devoted to schoolwork and lead students to develop better study habits.

Nonexperimental research on the effectiveness of post-secondary academic support services is mixed and, in many cases, largely descriptive in nature, with little in the way of convincing evidence of causal effects (see, e.g., surveys by Thomas R. Bailey and Marianna Alfonso 2005; Ernest T. Pascarella and Patrick T. Terenzini 1991; Veronica A. Lotkowski, Steven B. Robbins, and Richard J. Noeth 2004, and Susan C. Wyckoff 1998). On the other hand, more rigorous studies using experimental and quasi-experimental designs, mostly for high school students, paint a more promising picture (e.g., Jean Baldwin Grossman and Joseph P. Tierney 1998, Victor Lavy and Analia Schlosser 2005, Eric P. Bettinger and Bridget Terry Long 2005, and Dan Bloom and Colleen Sommo 2005).

The evidence on the impact of post-secondary merit scholarships is more limited than that for academic support services, though interest in this topic is growing as scholarship programs have expanded. A number of quasi-experimental evaluations suggest programs like Georgia HOPE boost college attendance and completion (Dynarski 2002, 2005; Christopher Cornwell, David B. Mustard, and Deepa J. Sridhar 2006). A few recent studies look at the impact of financial incentives on the performance of college students, though some of this research is still on-going (e.g.,

[1] The National Merit program awards roughly 8,200 scholarships to students selected from 1.4 million PSAT-takers.

Stephen L. DesJardins and Brian P. McCall 2006, and Thomas Brock and Lashawn Richburg-Hayes 2006).[2] An evaluation of gradually increasing tuition payments in response to delayed completion by Pietro Garibaldi et al. (2007) found substantial effects on Italian college women, while Edwin Leuven, Hessel Oosterbeek, and Bas van der Klaauw (2003) report mixed effects in a study of modest financial incentives in Amsterdam.

This paper reports on the Student Achievement and Retention Project (Project STAR), a randomized evaluation of academic services and incentives at one of the satellite campuses of a large Canadian university. In American terms, this institution can be thought of as a large state school, with tuition heavily subsidized. Most students are from the local area and have a common secondary school background. For the purposes of the study, all first-year students entering in September 2005, except those with a high school grade point average (GPA) in the upper quartile, were randomly assigned to one of three treatment groups or a control group. One treatment group was offered an array of support services including access to mentoring by upper-class students and supplemental instruction. A second group was offered substantial cash awards, up to the equivalent of a full year's tuition, for meeting a target GPA. A third treatment group was offered a combination of services and incentives, an intervention that has not been looked at previously. The control group was eligible for standard university support services but received nothing extra.

The first finding that comes out of Project STAR is that women use services much more than men. This is consistent with a range of evidence pointing to smoother college careers for women (see, e.g., Brian A. Jacob 2002). A second finding, also related to service use, is the appearance of a strong interaction between the offer of fellowships and service take-up. Students in the combined group were much more likely to use services than those offered services with no opportunity to win fellowships. Incentives therefore had the immediate short-term effect of increasing the rate at which students sought academic support.

The effects of the STAR intervention on students' academic performance are more mixed than the effects on use of services. Male achievement was essentially unchanged by the intervention, a disappointment for those who look to incentives to be an easy fix for boys' academic performance and completion problems. On the other hand, we find reasonably clear evidence of a sizeable impact on females.

The effects on women are strongest for those offered both fellowships and services (i.e., the combined group). Although women in both the combined and the fellowship-only treatment groups had markedly better fall-term grades, first-year GPAs were significantly higher only in the combined group. The combined group also earned more credits and had a significantly lower rate of academic probation at year's end. Importantly, women in the combined group continued to outperform

[2] Other evidence on incentives for academic performance comes from pre-college populations. For example, Ashworth et al. (2001) discuss a nonexperimental evaluation of stipends for high school students who stay in school, and Michael Kremer, Edward Miguel, and Rebecca Thornton (2004) report results from a randomized evaluation of a merit scholarship program for adolescent girls in Kenya. Angrist and Lavy (2002) evaluate a demonstration program that provided substantial cash incentives to high school students in Israel. Angrist et al. (2002) evaluate the impact of school vouchers in Colombia that required students at least meet grade promotion standards for eligibility. These studies consistently uncover positive effects for some types of primary or secondary school students.

the rest of the STAR population in the second year. This is in spite of the fact that fellowships and services were available in the first year only. These results suggest that students in the combined group benefited from a lasting improvement in study habits, a result we interpret as increased human capital, though not necessarily in the form of academic subject knowledge.

II. The Student Achievement and Retention (STAR) Demonstration Project

A. *Study Design*

The STAR demonstration randomly assigned entering first-year undergraduates to one of three treatment arms: a service strategy known as the Student Support Program (SSP), an incentive strategy known as the Student Fellowship Program (SFP), and an intervention offering both known as the SFSP. The SSP offered 250 students access to both a peer-advising service and a supplemental instruction service in the form of Facilitated Study Groups (FSGs). Peer advisors were trained upper-class students in the treated students' program of study. Advisors were meant to offer academic advice and suggestions for coping with the first year of school. Advisors e-mailed participants regularly and were available to meet at the STAR office. FSGs are class-specific sessions designed to improve students' study habits without focusing on specific course content. FSG facilitators were also trained upper-class students. The FSG model is widely used in North American colleges and universities (David R. Arendale 2001).

The SFP offered 250 students the opportunity to win merit scholarships for solid but not necessarily top grades in the first year. Award targets were set based on high school grade quartiles. Participants from the lowest grade quartile were offered \$5,000, roughly a year's tuition, for a B average (a GPA of 3.0) or \$1,000 for a C+ (a GPA of 2.3).[3] Participants from the second quartile were offered \$5,000 for a B+ or \$1,000 for a B−. Award thresholds were raised to A− and B for those in the third quartile. Upper-quartile students were not eligible to participate in the study. To qualify for a fellowship, SFP students had to take at least four courses per term and register to attend the second year of their program (a full load, required to complete a degree program in four years, is five courses per year).

A third treated group of 150 SFSP students was offered both the SSP and SFP. It is important to note, however, that other than having access to both services and scholarships, there was no link between the two strategies in this group. In particular, SFSP students need not have used SSP services to be eligible for a fellowship. Finally, the STAR demonstration included a control group of 1,006 students, with whom program operators had no contact beyond a baseline survey that went to all incoming freshmen.[4]

[3] Fellowship, scholarship, and bursary amounts are tax exempt in Canada. These income sources are not counted when determining financial aid grant eligibility but are counted when determining eligibility for loans. Amounts are in Canadian dollars and were worth roughly \$0.90 US at the time.

[4] Sixteen percent of the first-year population received a fellowship offer, and 26 percent were invited to participate in one of the three treatment programs. We received few inquiries from controls or other nonprogram students.

The SSP strategy was motivated in part by the view that retention is strongly influenced by a student's interaction with other people who take an interest in their welfare (Wesley R. Habley 2004). Several universities match first year students with upper-class peers or faculty advisors who provide academic support. Susan C. Wyckoff (1998) suggests these informal and formal interactions increase the likelihood students stay in school. Few colleges, however, offer as extensive a mentoring program as the SSP component of STAR. Peer advisors in the STAR program had exceptional social and academic skills. They participated in a three-day training course and received continuous training and feedback from supervisors. Advisors e-mailed advisees at least biweekly as a reminder of the advisors' availability and to solicit questions about university assimilation, scheduling, studying, and time management. The advisors complemented existing student services by reminding advisees of the availability of STAR and non-STAR services, and by encouraging advisees to use these services and to attend tutorials and make use of faculty office hours. Advisors were also trained to identify circumstances that called for more professional help and to make appropriate referrals.

The second component of the SSP consisted of Facilitated Study Groups (FSGs). FSGs were voluntary, course-focused, weekly sessions open to all treated students in the SFP or SFSP. FSG student facilitators were previously successful in the course they were hired to facilitate. They attended the same course with their assigned STAR students and tried to help students develop reasoning skills useful for the subject they were facilitating. FSGs were designed to complement the regular content-based tutorials taught by graduate students. Rather than walking through sample problems, FSGs focus on critical thinking, note-taking, graphic organization, questioning techniques, vocabulary acquisition, and test prediction and preparation. FSGs are a type of supplemental instruction commonly used in North American universities (Lotkowski, Robbins, and Noeth 2004). A number of studies suggest students who participate in FSG-style supplemental instruction outperform nonparticipating peers (Dennis H. Congos and N. Schoeps 2003, Kari A. Hensen and Mack Shelley 2003, Peggy Ogden et al. 2003). The STAR demonstration offered FSGs for approximately half of the largest first-year courses.[5]

SFP grade targets were based on a trade-off between program costs and award accessibility. A high GPA target is, of course, less costly, but few low-skilled students are likely to qualify. A low GPA target may be expensive and probably has little effect on those who can easily meet the target.[6] Grade targets were therefore set as a function of high school GPA. Students in the top GPA quartile were dropped from the STAR sample because few in this group fail to graduate (7.2 percent of incoming students in 1999 in the top high school grade quartile had not graduated by 2006 compared to 35.3 percent of students in the other quartiles). For students in remaining quartiles, the $5,000 target was set so that without the intervention, about 5–10 percent would qualify based on historical data. The $1,000 target was

[5] FSGs were offered to treated students taking calculus (first year mathematics), computer science, biology, English, anthropology, management and commerce, political science, and philosophy. Some of the other large courses offered FSGs to all students because these services were already in place before the experiment began.

[6] Dynarski (2005) and Cornwell, Mustard, and Sridhar (2006) estimate that the vast majority of Georgia HOPE scholarship recipients would have maintained the first-year target GPA of 2.0 even in absence of the program.

set so that about 20–25 percent were expected to qualify in the absence of a treatment effect. For a subset of SFP and SFSP students, there was also an intermediate target of $2,500. The resulting GPA targets were between 2.3 (C+) and 3.0 (B) for the $1,000 award and between 3.0 (B) and 3.7 (A−) for the $5,000 award.[7] The full set of STAR GPA targets appears in a chart in the Appendix.[8]

Students receive one credit unit for taking a two-semester (fall and spring) course and half a credit unit for taking a one-semester (fall or spring) course. A full course load of five credits per year is typically required to finish an undergraduate degree program in four years. About 40 percent of students take a full course load in the fall and spring terms, but many who drop below the full course load also take courses over the summer. To give some students with fewer than five credits a shot at a merit scholarship while minimizing the incentive to take fewer courses, award eligibility was based on a student's top four credits over the fall and spring terms.

In addition to meeting grade targets, SFP and SFSP students were required to enroll for a second year (at any college) to be eligible for a fellowship. Fellowship checks were sent to students in August after students registered for their second year of college. All students with grades above their targets continued into the second year, without interruption and without changing university.

Shortly after they acknowledged receipt of program details, students in the SSP and SFSP were assigned advisors. The advisors e-mailed participants in an effort to set up an initial meeting. FSG times and locations were announced often. After the first semester, bookstore gift certificates (worth up to $50) were offered to those who attended FSGs or met with peer advisors. Wallet-sized reminder cards were mailed to SFP and SFSP participants in November detailing grade targets. A second reminder went out in February and a third in March.

B. *Student and School Background*

Table 1 reports means and differences in means by treatment group for key administrative and background variables. Almost all of the 1,656 full-time, first-year students selected for random assignment in August of 2005 registered for class in the fall. The 85 students who did not register by the start of the school year or were registered for no more than two courses on November 1 (a group we call "no-shows") were dropped from the main analysis. With or without the no-shows, there are no significant differences by treatment status in students' sex, age, last year of high school GPA, or mother tongue.[9] In July, prior to treatment selection, we surveyed all incoming first-year students. More than 90 percent of the 1,571 who registered for at least two courses completed this survey. The likelihood of completing the survey appears to be unrelated to treatment status.

[7] Treated students were not told how their GPA target was chosen. In response to one student inquiry, program operators explained that the targets were set individually for research purposes.

[8] The average grade for control students fell 3 percentage points relative to students in the same high school GPA quartile from the previous two years. Hence, award rates were lower than anticipated.

[9] Only 1 of the 31 no-shows in a treatment group signed up. The estimates in Table 1 suggest no-show probabilities are unrelated to treatment. The estimated treatment effects are essentially unchanged when no-shows are included.

Table 1—Descriptive Statistics

	Control mean (1)	Contrasts by treatment status: SSP vs. control (2)	SFP vs. control (3)	SFSP vs. control (4)	F-stat (all=control) (5)	Obs. (6)
Administrative variables						
Courses enrolled as of fall 2005	4.745 {1.370}	−0.053 [0.095]	0.015 [0.095]	−0.158 [0.118]	0.702 (0.551)	1,656
No show	0.054	0.002 [0.016]	−0.030 [0.016]*	0.020 [0.019]	1.852 (0.136)	1,656
Completed survey	0.898	−0.018 [0.022]	−0.010 [0.022]	−0.051 [0.028]*	1.228 (0.298)	1,656
Student background variables						
Female	0.574	−0.006 [0.036]	0.029 [0.035]	−0.005 [0.045]	0.272 (0.845)	1,571
High school GPA	78.657 {4.220}	0.170 [0.308]	0.238 [0.304]	−0.018 [0.384]	0.276 (0.843)	1,571
Age	18.291 {0.616}	−0.054 [0.045]	−0.033 [0.044]	0.026 [0.056]	0.752 (0.521)	1,571
Mother tongue is English	0.700	0.017 [0.033]	0.009 [0.033]	0.049 [0.041]	0.495 (0.686)	1,571
Survey response variables						
Lives at home	0.811	−0.040 [0.030]	0.009 [0.030]	−0.004 [0.038]	0.685 (0.561)	1,431
At first choice school	0.243	0.024 [0.034]	0.060 [0.033]*	0.047 [0.042]	1.362 (0.253)	1,430
Plans to work while in school	0.777	0.031 [0.032]	−0.066 [0.031]**	0.037 [0.040]	2.541 (0.055)	1,431
Mother a high school graduate	0.868	0.015 [0.026]	−0.021 [0.026]	−0.045 [0.033]	1.040 (0.374)	1,431
Mother a college graduate	0.358	0.053 [0.037]	−0.020 [0.036]	−0.052 [0.046]	1.487 (0.216)	1,431
Father a high school graduate	0.839	0.025 [0.028]	0.008 [0.027]	−0.017 [0.035]	0.416 (0.741)	1,431
Father a college graduate	0.451	0.021 [0.038]	−0.001 [0.037]	−0.024 [0.048]	0.216 (0.885)	1,431
Rarely puts off studying for tests	0.208	0.031 [0.032]	0.031 [0.031]	0.107 [0.040]***	2.534 (0.055)	1,431
Never puts off studying for tests	0.056	−0.019 [0.016]	−0.016 [0.016]	−0.032 [0.021]	1.206 (0.306)	1,431
Wants more than a BA	0.556	0.052 [0.038]	−0.029 [0.037]	0.073 [0.048]	(1.752) (0.155)	1,431
Intends to finish in 4 years	0.821	−0.008 [0.030]	−0.006 [0.029]	−0.063 [0.037]*	(0.942) (0.419)	1,431

Notes: Standard deviations are shown in braces in column 1. Standard errors are reported in brackets in columns 2–4. p-values for F-tests are reported in parentheses in column 5. The last column shows the number of nonmissing observations.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

The university in which this study was carried out is primarily a commuter school. Roughly 80 percent of students in our sample were living at home. Slightly less than a quarter identified this campus as their first choice for college. The majority planned to work at least part time while in college (and most worked in high school). Many of the students are immigrants or children of immigrants, as suggested by the fact that 30 percent have a non-English mother tongue.[10] The students' parents, however, are reasonably well-educated. Many have college degrees. Only about a quarter of the students claim to never or rarely procrastinate. On the other hand, 56 percent said they wanted more education than a bachelor's degree and 82 percent said they intended to complete their undergraduate program in four years. Among those who entered college in 2001, however, only 38 percent completed a degree this quickly. In this earlier cohort, the six-year graduation rate was about 70 percent, and 13 percent dropped out after the first year.

Merit scholarship programs like STAR may affect course enrollment decisions and/or the selection of courses by treated students. In view of this concern, Table 2 reports treatment effects on students' completed course load and the number of math and science credits completed (these courses are considered more difficult). The estimates reported in the table are coefficients on dummies for each of the three STAR treatment groups, estimated either with a small set of controls from administrative data or a larger set of controls that includes variables from the background survey.[11] For the most part, there is little evidence of a change in the number or type of courses for which students registered. An exception is the SFP effect on the number of math and science credits completed by men. It seems unlikely, however, that this positive effect on course difficulty is a response to the treatment (since financial incentives should lead students to shift to an easier load).[12]

A second selection issue addressed in Table 2 is the likelihood of having fall grades. This is important because some students take no one-semester courses and are therefore omitted from the sample used to analyze the impact on fall grades. As can be seen in the last two columns in Table 2, about 89 percent of those registered for at least two courses have a fall grade. The likelihood of being in this group is unrelated to treatment assignment.

[10] Few students are French-speaking. Most nonnative English speakers in our sample are from South or East Asia.

[11] Columns labeled "Basic controls" report estimates of the coefficient on assignment-group dummies in models that control for sex, mother tongue, high school grade quartile, and number of courses as of November 1. These variables come from administrative data. Columns labeled "All controls" include dummy variables for nine parental education categories (high school with and without degree, community college with and without certificate, university with and without bachelors degree, masters degree, doctor or professional degree, and don't know), and four categories for self reported procrastination (usually, often, occasionally, rarely, never). These variables were selected on the basis of their predictive power in the grades regressions discussed below.

[12] Students interviewed in focus groups (discussed in Section IV) said that the fellowship program did not influence their course selection. Two first-year instructors in economics and biology also reported that no student mentioned Project STAR when asking for a higher grade. Angrist et al. (2008) provide additional evidence that STAR had little or no affect on credits attempted and course difficulty (with difficult courses defined as those where 35 percent or more of the previous year's class received a grade no better than a D).

Table 2—Selection Effects

	Number of credits attempted		Number of math and science credits attempted		Has fall grades	
	Basic (1)	All controls (2)	Basic (3)	All controls (4)	Basic (5)	All controls (6)
Panel A: All						
Control group mean	4.049 (0.893)		1.095 (1.206)		0.893 (0.309)	
Offered SSP	0.076	0.049	0.073	0.081	0.004	0.000
	[0.056]	[0.060]	[0.085]	[0.091]	[0.022]	[0.023]
Offered SFP	0.020	0.037	0.096	0.115	0.007	−0.004
	[0.056]	[0.054]	[0.080]	[0.086]	[0.022]	[0.023]
Offered SSP and SFP	−0.070	−0.086	−0.148	−0.139	−0.039	−0.041
	[0.074]	[0.081]	[0.094]	[0.104]	[0.032]	[0.034]
Observations	1,571	1,431	1,571	1,431	1,571	1,431
Panel B: Men						
Control group mean	3.964 (0.944)		1.159 (1.240)		0.914 (0.281)	
Offered SSP	0.080	0.032	0.012	0.020	0.016	0.016
	[0.088]	[0.101]	[0.126]	[0.139]	[0.029]	[0.031]
Offered SFP	−0.123	−0.063	0.276	0.344	−0.035	−0.051
	[0.098]	[0.089]	[0.138]**	[0.149]**	[0.036]	[0.039]
Offered SSP and SFP	−0.133	−0.115	−0.117	−0.048	−0.068	−0.089
	[0.117]	[0.137]	[0.160]	[0.184]	[0.048]	[0.057]
Observations	665	594	665	594	665	594
Panel C: Women						
Control group mean	4.112 (0.848)		1.047 (1.179)		0.877 (0.328)	
Offered SSP	0.072	0.058	0.118	0.132	−0.003	−0.013
	[0.072]	[0.077]	[0.116]	[0.123]	[0.031]	[0.032]
Offered SFP	0.111	0.093	−0.015	−0.004	0.034	0.021
	[0.066]*	[0.071]	[0.096]	[0.104]	[0.027]	[0.029]
Offered SSP and SFP	−0.046	−0.108	−0.179	−0.153	−0.015	−0.011
	[0.097]	[0.103]	[0.116]	[0.127]	[0.042]	[0.044]
Observations	906	837	906	837	906	837

Notes: The table reports regression estimates of treatment effects on the dependent variables indicated at left. Robust standard errors are reported in brackets. The sample is limited to students registered for at least two courses as of November 1 with data on the relevant set of controls. "Basic controls" include sex, mother tongue, high school grade quartile, and number of credits enrolled. "All controls" includes basic controls plus responses to survey questions on procrastination and parents' education.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

C. *Consent Rates and Service Use*

Students randomly assigned to STAR treatment groups were asked to provide consent by signing up. Those who did not sign up were ineligible for services and fellowships. Signing up imposed no burden or obligation on program participants beyond the receipt of reminder e-mails and mailings including a biweekly e-mail from peer advisors in the service programs. Students assigned to the control group were not contacted. While all students initially selected were tracked with administrative data, sign-up among the treated serves as an indicator of student awareness and interest. A little over half of those randomly assigned to receive services in the SSP responded to the invitation to sign up, a statistic reported in panel A of Table 3 (columns 1 and 2 show estimates from regression models for sign up with basic and all controls). Consent rates were much higher for the SFP than for the SSP, about 87 percent versus 55 percent. SFSP consent rates were about 79 percent.

Women in each of the three treatment groups were much more likely than men to sign up to participate in STAR. For example, column 2 in panel B of Table 3 shows that 46 percent of men offered the SSP consented, in contrast with 61 percent of women, a statistic reported in the same column in panel C. Most students offered the SFP signed up, but a gap by sex remains, with 91 percent of women and 81 percent of men having signed up. Similarly, when offered both services and fellowships in the SFSP, 84 percent of women and 71 percent of men signed up.

The pattern of service use shows differences by treatment group and sex similar to those observed in sign-up rates. This can be seen in columns 3–4 of Table 3. In particular, service use was higher for those assigned to the SFSP (combined services and incentives) than for those assigned to the SSP (services alone).[13] For example, 26 percent of students offered services in the SSP either attended an FSG or met or e-mailed their advisor, while service use was close to 43 percent for students offered both services and incentives in the SFSP. Women were also much more likely to use services than men. SFSP service use by men was 29 percent, while 53 percent of women in the SFSP used services. The fact that service use was higher in the SFSP than in the SSP suggests that the opportunity to win a fellowship motivated students to use services.

Use rates for specific services are reported in columns 5–8 in Table 3. Students took advantage of the peer advising service more than the supplemental instruction offered through FSGs. About 12 percent of the SSP group attended at least one FSG (most of those who attended at least once attended more than once), while 15 percent of men and 26 percent of women met or e-mailed a peer advisor (excluding advisor-initiated contact). Usage rates for both types of services were higher in the SFSP than the SSP, with 49 percent of women in the SFSP having contacted a peer advisor and 16 percent having attended an FSG.

Take-up rates for the FSG services were lower than the rates we aspired to, and probably diluted somewhat by our inability to offer FSGs in every course in which STAR participants were enrolled (though 86 percent of subjects attended at least one course incorporating an FSG). Take-up was probably also reduced by the fact that we offered services to individual students as opposed to entire classrooms and by unavoidable scheduling conflicts. On the other hand, treated students made considerable use of the advising services. In our follow-up focus groups, participants indicated that they found peer advisors to be a valuable resource.

[13] Service use was about the same in both semesters.

Table 3—Program Sign-up and Use of Services

	Signed up for STAR		Received SSP services		Met with/emailed an advisor		Attended FSGs	
	Basic controls (1)	All controls (2)	Basic controls (3)	All controls (4)	Basic controls (5)	All controls (6)	Basic controls (7)	All controls (8)
Panel A: All								
Offered SSP	0.519	0.549	0.238	0.255	0.204	0.217	0.106	0.118
	[0.032]***	[0.034]***	[0.028]***	[0.029]***	[0.026]***	[0.028]***	[0.020]***	[0.021]***
Offered SFP	0.863	0.867						
	[0.022]***	[0.022]***						
Offered SSP and SFP	0.762	0.792	0.412	0.431	0.383	0.397	0.131	0.139
	[0.036]***	[0.036]***	[0.041]***	[0.044]***	[0.041]***	[0.043]***	[0.029]***	[0.031]***
Observations	1,571	1,431	1,571	1,431	1,571	1,431	1,571	1,431
Panel B: Men								
Offered SSP	0.447	0.464	0.194	0.206	0.145	0.149	0.096	0.107
	[0.049]***	[0.052]***	[0.039]***	[0.042]***	[0.035]***	[0.038]***	[0.029]***	[0.032]***
Offered SFP	0.792	0.806						
	[0.040]***	[0.040]***						
Offered SSP and SFP	0.705	0.708	0.298	0.291	0.282	0.270	0.115	0.112
	[0.058]***	[0.065]***	[0.058]***	[0.063]***	[0.057]***	[0.061]***	[0.042]***	[0.046]**
Observations	665	594	665	594	665	594	665	594
Panel C: Women								
Offered SSP	0.571	0.605	0.273	0.287	0.251	0.264	0.113	0.124
	[0.043]***	[0.044]***	[0.038]***	[0.040]***	[0.037]***	[0.040]***	[0.027]***	[0.029]***
Offered SFP	0.912	0.908						
	[0.024]***	[0.026]***						
Offered SSP and SFP	0.800	0.835	0.506	0.532	0.466	0.489	0.146	0.155
	[0.046]***	[0.043]***	[0.056]***	[0.058]***	[0.056]***	[0.058]***	[0.040]***	[0.042]***
Observations	906	837	906	837	906	837	906	837

Notes: The table reports regression estimates of treatment effects on the dependent variables indicated in column headings. Robust standard errors are reported in brackets. The sample is limited to students registered for at least two courses as of November 1 with data on the relevant set of controls. “Basic controls” include sex, mother tongue, high school grade quartile, and number of credits enrolled. “All controls” includes basic controls plus responses to survey questions on procrastination and parents education.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

D. *Student and Advisor Gender Interactions*

After signing up, SSP and SFSP students were assigned peer advisors. Assignment was based on common fields of study. There were 21 female advisors and 8 male advisors, so both women and men were more likely to be matched with women. Same-sex advisor matching may explain the greater use of advisors for females. Table 4 compares advisor communication by gender for the sample of 215 STAR participants who signed up for the SSP or SFSP.

The first column of Table 4 shows that women met with advisors more often than men did, a gap of 16 percentage points. The results in column 3, however, suggest that women were substantially less likely to meet with male advisors, while men were less likely to meet with female advisors. Students were also more likely to e-mail a same-sex advisor. These results point to a role for gender matching in the use of academic services. On the other hand, gender differences in sign up cannot be explained by the prevalence of female advisors since advisors were not assigned

Table 4—Analysis of Gender-Adviser Interaction

	By advisee gender			By advisor/advisee mix	
	Met with advisor (1)	Contacted advisor (including email) (2)		Met with advisor (3)	Contacted advisor (including email) (4)
Women [135]	0.385	0.496	Women with male advisors [31]	0.226	0.420
			Women with female advisors [104]	0.437	0.519
Men [80]	0.225	0.350	Men with female advisors [54]	0.203	0.334
			Men with male advisors [26]	0.269	0.385

Notes: The table shows contact rates by student and advisor gender. The sample includes all SSP and SFSP students who signed up for Project STAR. Sample counts by gender or gender mix appear in brackets. The overall sample size is 215.

until after students signed up. Moreover, gender differences in the use of services remain for same-sex advisor pairs.

E. *Evaluation Framework*

The bulk of the estimates reported below are intention-to-treat effects that make no adjustment for sign up. In cases where program effects are zero, a zero intention-to-treat effect implies a zero effect on participants. More generally, however, intention-to-treat effects are diluted by noncompliance. For example, some of those offered the fellowship program were ineligible for fellowships because they did not sign up. Likewise, students who did not sign up for the SSP and SFSP could not use services and were not contacted by peer advisors. This reduces the overall impact of the offer of treatment. The empirical section therefore concludes with a set of estimates that use the offer of services as an instrumental variable (IV) for program participation (i.e., sign-up). This generates an estimate of the effect of treatment on those who signed up to participate.

The IV adjustment works as follows. Let P_i denote participants (in this case, those who gave consent), and let Z_i denote the randomly assigned offer of treatment. The IV formula in this simple case is the adjustment to intention-to-treat effects originally proposed by Howard S. Bloom (1984):

$$E[Y_{1i} - Y_{0i} | P_i = 1] = \{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]\} \div \Pr[P_i = 1 | Z_i = 1]. \tag{1}$$

This is the intention-to-treat effect divided by the compliance rate in the treatment group. A regression-adjusted estimate of the effect on program participants can be constructed using two-stage least squares (2SLS), where Z_i acts as an instrument for P_i. The result is a covariate-weighted average effect of treatment on the treated (Guido W. Imbens and Angrist 1994). In the SSP and SFSP, a further distinction can be made between compliance via sign up and compliance via service use. But the availability of services and the weekly e-mails sent by peer advisors is an intervention to which all SSP and SFSP participants who signed up were exposed whether or not they actively sought services. In focus groups, treated students reported that they took note of the advisors' e-mails even if they did not respond. We therefore make no adjustment for the difference between sign-up and usage in the 2SLS analysis.

III. Results

A. *Main Results*

Our analysis of achievement effects begins by looking at students' average grades in the fall semester and at their official GPAs at the end of the first year of study. The fall grade variable is a credit-weighted average on a 0–100 grading scale for those who took one or more one-semester courses. This variable provides an initial measure of program impact. Although some students (about 11 percent) are omitted from the fall grades sample because they took no one-semester courses, membership in the fall grades sample appears to be unrelated to treatment status (as shown in Table 2). The first-year GPA variable is the registrar's official end-of-year grade point average, computed on a scale of 0–4. For example, a GPA of 3.3 denotes a B+. This is the variable according to which STAR fellowships were awarded.

Students assigned to the SFP earned fall grades about 1.8 percentage points higher than students in the control group, while those assigned to the SFSP earned grades 2.7 points higher than controls. Both of these effects are significantly different from zero, as can be seen in the first column of panel A in Table 5, which reports treatment effects estimated in the pooled sample of men and women. The results reported in this and subsequent tables are from regressions including the same set of "all controls" used to construct the estimates in Tables 2 and 3.[14]

Because both of the fellowship treatments produced substantial and significant effects on fall grades, we report estimates from models that pool these into a single "any-SFP" effect. Models with a single dummy indicating assignment to either the SFP or the SFSP generate a combined any-SFP effect of 2.1 (s.e. = 0.73) reported in column 4. The fall grades estimates can be compared to a standard deviation of about 12. In contrast with the significant estimates for the two fellowship groups, the corresponding SSP effect is small and insignificant, though estimated with approximately the same precision as the SFP and SFSP effects.

The overall impact of both fellowship treatments on fall grades is driven entirely by large and significant effects on women. This is apparent in columns 2 and 3 of Table 5. Women assigned to the SFP earned a fall grade 2.6 points higher than the control group, while women assigned to the SFSP earned a fall grade 4.2 points higher than controls. Thus, the estimates for women suggest the combination of services and fellowships offered in the SFSP had a larger impact than fellowships

[14] Results with basic controls are similar, though the estimates for males with the smaller set of controls show a few marginally significant negative effects. The pattern of these effects is not consistent across outcomes, treatment types, or years, and therefore seems likely to be due to chance. It is also worth noting that the treatment effects reported here differ somewhat from those in our initial December 2006 National Bureau of Economic Researach (NBER) working paper where we reported preliminary findings. In 2007, we received more complete and more accurate administrative records. These records were used to construct the estimates reported here, in our 2007 IZA Working paper, and in our final report to CMSF (Angrist et al. 2008).

Table 5—Treatment Effects on First Year Outcomes in the Sample with Fall Grades

	SFP by type			Any SFP		
	All (1)	Men (2)	Women (3)	All (4)	Men (5)	Women (6)
Panel A: Fall grade						
Control mean	64.225	65.935	62.958	64.225	65.935	62.958
	(11.902)	(11.340)	(12.160)	(11.902)	(11.340)	(12.160)
SSP	0.349	−0.027	0.737	0.344	−0.014	0.738
	[0.917]	[1.334]	[1.275]	[0.917]	[1.332]	[1.274]
SFP	1.824	0.331	2.602			
	[0.847]**	[1.233]	[1.176]**			
SFSP	2.702	−0.573	4.205			
	[1.124]**	[2.010]	[1.325]***			
SFP (any)				2.125	0.016	3.141
				[0.731]***	[1.164]	[0.972]***
Observations	1,255	526	729	1,255	526	729
Panel B: First year GPA						
Control mean	1.805	1.908	1.728	1.797	1.885	1.731
	(0.902)	(0.908)	(0.891)	(0.904)	(0.910)	(0.894)
SSP	0.073	0.011	0.116	0.071	0.008	0.116
	[0.066]	[0.107]	[0.082]	[0.066]	[0.107]	[0.082]
SFP	0.010	−0.110	0.086			
	[0.064]	[0.103]	[0.084]			
SFSP	0.210	0.084	0.267			
	[0.092]**	[0.162]	[0.117]**			
SFP (any)				0.079	−0.042	0.147
				[0.056]	[0.095]	[0.073]**
Observations	1,255	526	729	1,255	526	729

Notes: The table reports regression estimates of treatment effects on fall grades and first-year GPA computed using the full set of controls. Robust standard errors are reported in brackets. The sample is limited to students registered for at least two courses as of November 1 with data on the relevant set of controls and at least one fall grade. The last three columns report estimates from a model that combines the SFP and SFSP treatment groups into "SFP (any)."

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

alone. The average fellowship effect (i.e., combining the SFP and SFSP groups using a single any-SFP dummy) is 3.1 (s.e. = 0.97). In contrast with the results for women, the estimated effects of both fellowship treatments on men are much smaller, and none are significantly different from zero.

The first-year GPA results are weaker than the fall grades results, as can be seen in panel B of Table 5 (these results use the same sample as used for panel A). In particular, by the end of the first year, the SFP effects on women had faded. The estimated SFP effect on women's first-year GPA is 0.086 (s.e. = 0.084), not too far from the insignificant SSP effect of 0.12 (s.e. = 0.082). On the other hand, the effect of the SFSP treatment remains large and significant at about 0.27 (s.e. = 0.12) for women and 0.21 (s.e. = 0.092) overall. Again, the overall result is driven by the effect on women. In standard deviation units, the SFSP effect on GPAs for women is only a little smaller than the corresponding fall grades effect (0.3σ versus 0.35σ). Combining both the SFP and the SFSP estimates in a single any-SFP effect generates a significant GPA result for women of about 0.15 (s.e. = 0.073). In standard deviations units, this is markedly smaller than the corresponding any-SFP effect on fall grades because there is no SFP-only effect on GPA.

The first-year GPA effects in the full sample are similar to those in the sample with fall grades. The full-sample GPA results are reported in panel A of Table 6 separately for year one and year two. Similar to the corresponding estimate in Table 5, the SFSP effect on women's first-year GPA is 0.24 (s.e. = 0.11). Again, the estimated effects on men are small and insignificant, as is the estimated effect of SFP-only (for men this estimate is negative).

A striking finding that emerges from Table 6 is the persistence of SFSP program effects on the GPAs of women into the second year. This estimate, at 0.28 (s.e. = 0.11), differs only slightly from the first-year effects. Thus, the SFSP treatment appears to have generated a robust improvement in performance that extended beyond the one-year period in which fellowships were awarded. Consistent with the fact that fellowships alone did not lead to better outcomes, the improvement in second-year GPA suggests the SFSP intervention led to a lasting improvement in study habits or skills.

An alternative measure of student academic performance is academic probation. Students with an average GPA below 1.5 after attempting to complete 4 credits are placed on academic probation and are at risk of suspension.[15] Many students spend time on probation. For example, 22 percent of the control group in our study were on academic probation at the end of their first year.

The SFSP intervention appears to have reduced the proportion of students on probation at the end of their first year of school, a result that can be seen in panel B of Table 6. The overall reduction is 6.9 (s.e. = 3.6) percentage points in the combined sample of men and women. For women, the SFSP effect on the first-year probation rates is −0.10 (s.e. = 0.051), a substantial reduction. On the other hand, the probation rates for men in the SFSP group are only slightly (and insignificantly) lower than the probation rates in the control group. Like the GPA effects in panel A, treatment effects on women's probation rates continued in the second year. The second-year probation effect on SFSP women, reported in column 6 of Table 6, declines to −0.097 (s.e. = 0.047), with an insignificant −0.053 (s.e. = 0.038) effect overall.[16]

A potential problem with the probation outcome is the fact that a student must have attempted to complete at least four credits to be classified as being on probation. Many students take a lighter load and should not necessarily be seen as making good progress in their studies even though they are not on probation. Panel C of Table 6 therefore shows results for an alternative measure of academic standing, a "Good Standing" variable that identifies students who have attempted at least four credits and are not on probation. About 47 percent of first-year students were in good standing at the end of the year. Mirroring the probation results, women in the SFSP treatment group were substantially and significantly more likely than controls to be

[15] A credit is awarded for each year-long course successfully completed. A half of a credit is awarded for each one-semester course successfully completed.

[16] This second-year variable codes students who have withdrawn in the second year as on probation.

Table 6—Treatment Effects on First and Second Year Outcomes

	Year 1			Year 2		
	All (1)	Men (2)	Women (3)	All (4)	Men (5)	Women (6)
Panel A: GPA						
Control mean	1.794	1.871	1.739	2.040	2.084	2.008
	(0.915)	(0.904)	(0.920)	(0.884)	(0.901)	(0.871)
SSP	0.011	0.017	0.002	0.050	−0.021	0.090
	[0.063]	[0.102]	[0.080]	[0.074]	[0.121]	[0.092]
SFP	−0.040	−0.144	0.038	−0.018	−0.081	0.030
	[0.061]	[0.098]	[0.080]	[0.066]	[0.108]	[0.085]
SFSP	0.168	0.016	0.244	0.072	−0.170	0.276
	[0.086]*	[0.146]	[0.111]**	[0.091]	[0.161]	[0.106]***
Observations	1,399	577	822	1,241	521	720
Panel B: On Probation/withdrew						
Control mean	0.221	0.159	0.266	0.247	0.253	0.243
	(0.415)	(0.366)	(0.443)	(0.431)	(0.435)	(0.429)
SSP	−0.015	0.016	−0.039	0.017	0.002	0.044
	[0.031]	[0.045]	[0.043]	[0.033]	[0.051]	[0.045]
SFP	−0.021	0.012	−0.055	0.008	0.011	−0.005
	[0.031]	[0.049]	[0.041]	[0.032]	[0.053]	[0.041]
SFSP	−0.069	−0.022	−0.100	−0.053	0.014	−0.097
	[0.036]*	[0.055]	[0.051]**	[0.038]	[0.066]	[0.047]**
Observations	1,418	590	828	1,418	590	828
Panel C: Good standing						
Control mean	0.466	0.486	0.451	0.633	0.643	0.626
	(0.499)	(0.500)	(0.498)	(0.482)	(0.480)	(0.484)
SSP	0.042	−0.058	0.104	−0.023	−0.019	−0.032
	[0.035]	[0.055]	[0.046]**	[0.036]	[0.056]	[0.048]
SFP	0.021	−0.041	0.071	0.012	0.000	0.035
	[0.035]	[0.056]	[0.047]	[0.035]	[0.059]	[0.044]
SFSP	0.062	−0.023	0.108	0.085	0.020	0.131
	[0.048]	[0.077]	[0.065]*	[0.043]**	[0.071]	[0.055]**
Observations	1,418	590	828	1,418	590	828
Panel D: Credits earned						
Control mean	2.363	2.453	2.298	2.492	2.468	2.509
	(0.986)	(1.069)	(0.917)	(1.502)	(1.525)	(1.486)
SSP	0.054	−0.066	0.130	−0.098	−0.176	−0.070
	[0.073]	[0.109]	[0.101]	[0.115]	[0.175]	[0.153]
SFP	−0.012	−0.157	0.084	0.027	0.155	−0.024
	[0.064]	[0.106]	[0.082]	[0.108]	[0.180]	[0.137]
SFSP	0.092	−0.196	0.269	0.072	−0.240	0.280
	[0.087]	[0.150]	[0.108]**	[0.130]	[0.206]	[0.172]
Observations	1,418	590	828	1,418	590	828

Notes: The table reports regression estimates of treatment effects computed using the full set of controls. Robust standard errors are reported in brackets. The sample is limited to students registered for at least two courses as of November 1, with data on the relevant set of controls. The GPA outcome samples include students with a GPA for each year. The probation variable indicates academic probation in the first year and probation or withdrawal in the second year. The credits earned and good standing variables are zero in the second year for those who withdrew. SFSP treatment effects for men and women are significantly different at the 5 percent level or better for year two GPA, credits earned in year one, and credit-weighted GPA in both years. The sex difference in effects on year two credits earned is significant at the 10 percent level.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

in good standing at the end of the first and second years. The SSP treatment group was also more likely than the control group to be in good standing, with an effect the same size as for the SFSP group. In contrast with the SFSP results, however, the SSP effect on good standing is not corroborated by effects on GPA or probation.

Panel D in Table 6 shows estimated effects on total credits earned, with a zero recorded for students who had withdrawn by the second year (effects on withdrawal are not significant). In addition to benefiting from an increase in grades, the results indicate that women in the SFSP treatment group earned more credits than those in the control group. The estimated treatment effect on credits is 0.27 (s.e. = 0.11) in the first year and 0.28 (s.e. = 0.17) in the second year. The credits earned and academic standing results are encouraging. They suggest that, for women at least, the SFSP treatment achieved a hoped-for result—an improvement in the rate at which students progress through their studies.[17]

B. *Effects on the Grade Distribution*

The results in Table 6 suggest the STAR intervention had an impact on female students who were at risk of probation or loss of academic standing. These relatively low-achievers were not very likely to win fellowships, raising the question of whether all of the students affected were in this low achieving group. To investigate this further, Figures 1A and 1B plot the GPA distribution for each of the STAR treatment groups along with the GPA distribution in the control group separately for men and women. The GPA variable in these plots is adjusted for high school GPA to produce the same award thresholds for all students.[18]

Consistent with the estimates in Tables 5 and 6, Figure 1A offers little evidence of a systematic shift in any of the grade distributions for men. Among women, the clearest evidence of a shift appears for the SFSP group, with a distribution moved to the right almost everywhere except for a segment where the treatment and control densities overlap. The clearest distribution shift is in the lower-middle portion of the grade distribution. As in panels B and C of Table 6, the SFSP intervention appears to have reduced the number of women with GPAs low enough to put them on probation. Moreover, the GPA shift appears more marked in areas of the grade distribution below the STAR fellowship award thresholds (also indicated in the figure). This suggests the program might have been more effective with lower fellowship targets.[19]

[17] In view of the similarity of the GPA results across years, we also constructed more precise estimates of a common treatment effect by stacking first- and second-year observations (with standard errors clustered by student). Stacked estimation generated a significant treatment effect for women in the SFSP treatment group with 10–20 percent reduction in standard errors relative to single-year estimates. See Table 7 for an example. We also looked at a credit-weighted GPA outcome variable, constructed as GPA $\times$ (credit units completed/5). Consistent with the GPA and credits-earned outcomes, the SFSP had a strong positive effect on credit-weighted GPA for SFSP women.

[18] Specifically, the college GPA variable in this figure equals GPA $+$ 0.3 for those in the lowest high school GPA quartile and GPA $-$ 0.3 for those in the second highest high school GPA quartile.

[19] To further investigate the location of the STAR-induced distribution shift, we coded a dummy for theoretical fellowship eligibility in both the treatment and control groups and used this as the dependent variable in a regression on treatment dummies and covariates. Roughly 16 percent of all control students finished their first year with a GPA that qualified for a \$1,000 payment. The eligibility rates for students in the treatment groups were similar. There was an increase in the likelihood that women in both fellowship groups met the standard for a \$1,000 award, 0.071 in the SFSP group, but this difference is not significantly different from zero (s.e. = 0.47).

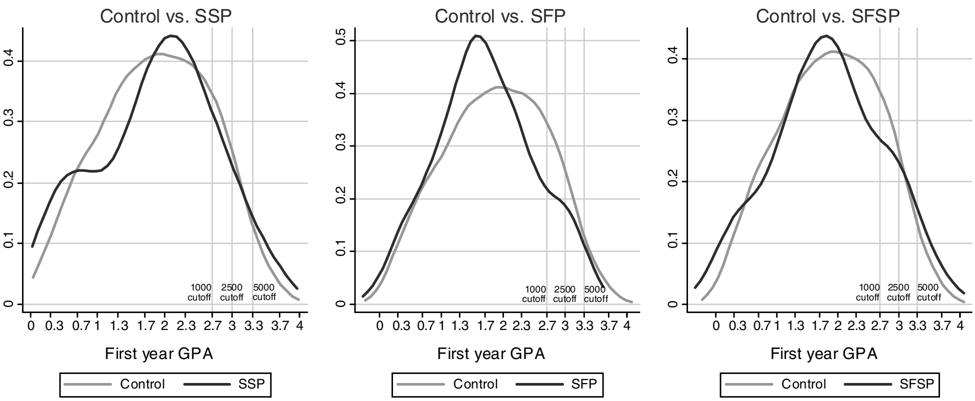

Figure 1a. Males' Normalized First-Year GPA

Note: These figures plot the smoothed kernel densities of first year GPA.

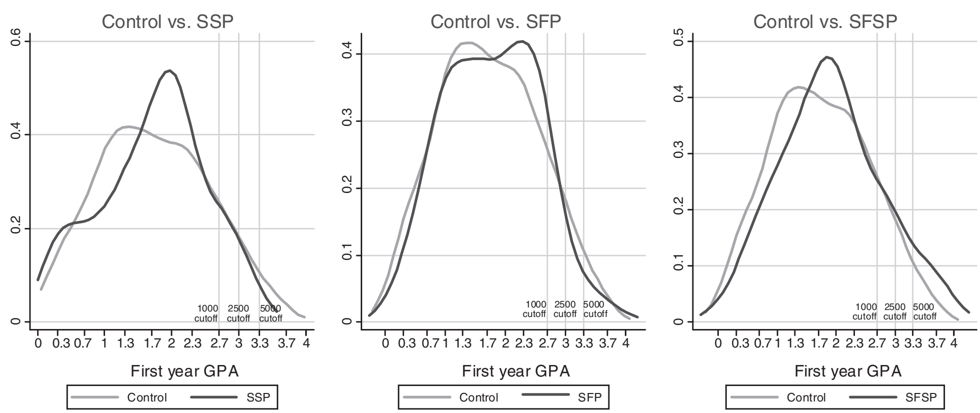

Figure 1b. Females' Normalized First-Year GPA

Note: These figures plot the smoothed kernel densities of first year GPA.

Treatment effects on the GPA distribution are further quantified in Table 7, which reports quantile regression estimates at the upper and lower deciles, upper and lower quartiles, and the median. To increase precision, and in view of the similarity of the results across years, this table was constructed using stacked data from year one and year two of the experiment. As a benchmark, the first column of the table reports OLS estimates using stacked data. These results are similar to those reported in Table 6, though somewhat more precise. For example, the SFSP effect for females in

There was a marginally significant 0.067 gain in $5,000 eligibility rates for SFSP women. The SSP does not appear to have affected fellowship eligibility. In all, 43 students from the SFP (10 males and 33 females) received a fellowship, as did 29 students from the SFSP (12 males and 17 females).

Table 7—OLS and Quantile Treatment Effects on GPA (*Stacked*)

	OLS	Quantile regressions				
		0.1	0.25	0.5	0.75	0.9
Quantile	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Males						
Control mean or quantile	1.972 (0.908)	0.750	1.340	2.040	2.660	3.140
SSP	−0.005 [0.096]	−0.075 [0.174]	0.038 [0.162]	0.107 [0.123]	0.039 [0.119]	−0.063 [0.134]
SFP	−0.108 [0.088]	−0.015 [0.159]	−0.045 [0.143]	−0.185 [0.1150]	−0.204 [0.128]	−0.084 [0.149]
SFSP	−0.078 [0.137]	−0.195 [0.221]	−0.211 [0.227]	−0.108 [0.209]	0.120 [0.192]	0.210 [0.154]
Observations	1,091					
Panel B: Females						
Control mean or quantile	1.865 (0.907)	0.630	1.225	1.900	2.530	3.080
SSP	0.043 [0.075]	−0.065 [0.143]	0.088 [0.137]	0.131 [0.09600]	0.088 [0.093]	−0.074 [0.092]
SFP	0.037 [0.073]	0.014 [0.101]	−0.044 [0.106]	0.003 [0.09800]	0.067 [0.0920]	−0.013 [0.099]
SFSP	0.264 [0.094]***	0.389 [0.150]***	0.271 [0.125]**	0.187 [0.125]	0.269 [0.1460]*	0.296 [0.1190]**
Observations	1,540					
Panel C: Female SFSP effects with limited sets of covariates						
Control for year and high school GPA quartile						
SFSP	0.283 [0.093]***	0.310 [0.1430]**	0.190 [0.141]	0.240 [0.101]**	0.250 [0.1580]	0.240 [0.1350]*
Control only for year						
SFSP	0.221 [0.1]**	0.280 [0.2060]	0.220 [0.146]	0.220 [0.104]**	0.190 [0.130]	0.010 [0.2260]
Observations	1,540					

Notes: Standard errors are reported in brackets; standard deviations are in parentheses in column 1. Standard errors are clustered at the student level. Quantile standard errors are bootstrapped using 500 replications. The sample stacks first- and second-year data on students registered for at least two classes as of November 1 who had valid values for the dependent variable and for the full set of controls. The models in panels A and B control for year and the full set of controls. The models in the first line of panel C control for year and high school GPA quartile. The models in the second line of panel C control only for year.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

the two-year sample is 0.264 with a standard error of 0.094, below both single-year standard errors of around 0.11.

The largest SFSP effect on the GPA distribution occurs at the 0.1 quantile, a finding that lines up with the GPA distribution shifts seen in the figures. But there is reasonably clear evidence of a shift at higher quantiles as well. In particular, effects at the 0.75 and 0.9 quantiles are close to the lower quartile impact of 0.271 (s.e. = 0.125). The smallest quantile effect is at the median, an estimate of 0.187 (s.e. = 0.125). Thus, the quantile regression estimates provide less evidence of a skewed impact than do the figures.

In this context, however, it's important to note that quantile regression estimates capture a conditional distribution shift: the effect of treatment with other included covariates held constant. The relatively flat quantile profile may mask overall asymmetry if treatment effects vary with included covariates.[20] In panel C of Table 7, we therefore report a set of SFSP effects using alternative sets of controls. When the only controls are high school GPA and a year dummy, the pattern of quantile effects is similar to that estimated with all controls. Interestingly, however, the quantile profile shows a relatively larger impact at the median and below when the high school GPA variable is dropped. In particular, the estimated effect at the 0.75 quantile becomes smaller without the high school GPA controls, while the significant effect at the 0.9 quantile becomes insignificant and close to zero. This suggests that the SFSP shifted the entire distribution of scores for students with low high school GPAs (a conditional location shift) with less of an effect on better students.

C. *Subgroup Effects*

The results above paint a reasonably clear picture of substantial SFSP achievement effects for women but not for men. The results for other subgroups, however, are not as clear cut. We estimated heterogeneous effects by high school GPA, expected hours of work, self-reported funding concerns, and parents' educational attainment. The additional analysis generated inconsistencies that may be due to chance variation. This fact, combined with a lack of strong prior beliefs to guide the subgroup inquiry, makes the analysis inconclusive. In the case of men and women, in contrast, there is considerable evidence from earlier evaluations that we might expect strong sex differentials in treatment effects. The pattern of estimates by sex is robustly consistent with this view.

The subgroup results are briefly summarized here. For a full report, see Angrist et al. (2008), available on request. For students with below median high school grades, SFSP effects on first-year GPA and credits earned are higher for those with a better high school record but about the same when looking at total credits earned. When stratified by expected employment, the estimated first-year GPA effects for SFSP women are higher among those expecting to work more but similar in the second year. There is little meaningful variation in treatment effects according to whether or not students were concerned about funding.

Perhaps the clearest pattern coming out of the subgroup analysis is that the SFSP effects are larger for women whose parents did not attend college. Specifically, the GPA and total credits earned results are significantly higher for first-college-generation SFSP women than for those whose parents had at least some college education. In fact, estimated treatment effects on first-year credits earned are also marginally significant for SFP and SSP women in the first-generation group. It should be noted, however, that we did not find parallel results in other breakdowns that are correlated

[20] The distinction between conditional and marginal quantile treatment effects is emphasized in recent work by Sergio Firpo (2007) and Markus Frölich and Blaise Melly (2007), who provide methods for the estimation of marginal treatment effects from conditional treatment effects. By virtue of random assignment, however, both the estimator proposed by Firpo (2007) and standard quantile regression without controls estimate the unconditional quantile treatment effect.

with socioeconomic status. The estimated effects of SSP and SFP are consistently zero across most cuts of the data.

D. *Two-Stage Least Squares Estimates*

Intention-to-treat effects are diluted by the fact that some treated students failed to sign up and were therefore ineligible to participate in STAR programs. We refer to students who signed up as participants. The average effect of treatment on participants provides a better indicator of the impact of treatment on the subpopulation that was directly affected by incentives and services. Moreover, in a future intervention of this sort, we might expect sign-up rates to be higher or simply extend services and incentives to all members of the freshman class. As a practical matter, effects on participants are larger than intention-to-treat effects, with the proportional increase equal to the reciprocal of the treatment-group-specific sign-up rate.

Estimates of the effect of each STAR treatment on participants were constructed using a model that parallels the intention-to-treat specifications. The first version of this model allows separate effects on participants in each program. In particular, we used studied first and second year data to estimate the following equation by 2SLS:

$$y_{it} = X_i'\delta_t + \alpha ssp_i^* + \beta sfp_i^* + \gamma sfsp_i^* + \xi_i + \varepsilon_{it}, \tag{2}$$

where ssp_i^*, sfp_i^*, and $sfsp_i^*$ indicate program participants (i.e., those who signed up). The participation variables were treated as endogenous and the three program-assignment dummies $(ssp_i, sfp_i, \text{ and } sfsp_i)$ used as instruments. This generalizes the compliance adjustment in (1) to include covariates and multiple treatments. [21]

In addition to estimates of equation (2), we constructed 2SLS estimates of a combined any-SFP effect by estimating the parameter δ in the equation

$$y_{it} = X_i'\delta_t + \alpha ssp_i^* + \delta(sfp_i^* + sfsp_i^*) + \xi_i + \varepsilon_{it}, \tag{3}$$

using the same instruments as were used to estimate equation (2). This model is over identified, a fact that motivates our investigation of it. In this context, the over-identification test statistic can be interpreted as a test for whether the 2SLS estimates of β and γ are equal in the just-identified model (Angrist 1991). In other words, we use the over-identified model to simultaneously estimate a common fellowship-related treatment effect and to test whether the model satisfies an exclusion restriction that implies that the primary causal force is fellowships per se, with or without the interaction with service. Although the reduced-form estimates weigh against this, 2SLS rescaling to take account of differences in take-up rates make this investigation of independent interest.

[21] The individual error component in equations (2) and (3), ξ_i, captures the year-to-year correlation in outcomes for a given student. The standard errors are clustered by student to account for this. Because the reduced form estimates for men are always zero, 2SLS estimates are reported for women only.

Table 8—2SLS Estimates for Women (*Stacked*)

	GPA		On probation/withdrew		Credits earned	
	SFP by type (1)	Any SFP (2)	SFP by type (3)	Any SFP (4)	SFP by type (5)	Any SFP (6)
Control mean	1.865 (0.907)		0.254 (0.436)		2.404 (1.239)	
SSP sign-up	0.071 [0.126]	0.068 [0.124]	0.004 [0.053]	0.005 [0.053]	0.049 [0.173]	0.046 [0.171]
SFP sign-up	0.041 [0.081]		−0.033 [0.033]		0.034 [0.104]	
SFSP sign-up	0.315 [0.112]***		−0.117 [0.043]***		0.327 [0.148]**	
SFP (any) sign-up		0.125 [0.070]*		−0.058 [0.029]**		0.121 [0.091]
Overid test $\chi^2(1)$		4.487 (0.034)		2.900 (0.089)		3.211 (0.073)
Observations	1,542	1,542	1,656	1,656	1,656	1,656

Notes: The table reports 2SLS estimates of the effect of signing up for the treatment indicated. Dummies for treatment assigned are used as instruments. The sample stacks first and second year data on women registered for at least two classes as of November 1 who had valid values for the dependent variable and the full set of controls. Standard errors adjusted for student clustering are reported in brackets. The models in columns 2, 4, and 6 pool the SFP and SFSP sign-up dummies. The chi-square statistic is the over-identification test statistic for these specifications, with p-values reported in parentheses.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

The impact of the SFSP on female participants GPAs, reported in column 1 of Table 8, is a little over one-third of a standard deviation, a substantial effect. The effects on probation and credits earned are a little smaller but also substantial. An important result in Table 8 is the precision of the estimated effects on participants. The 2SLS standard errors give us an idea of the size of the effects on participants we might expect to be able to detect. The estimated standard error for the SSP effect on GPA, 0.126, is of a magnitude that would allow us to detect an effect size on the order of one-quarter of a standard deviation. In terms of detectable effect size, the precision of the probation estimates is similar. Thus, in spite of lower take-up for the SSP, the 2SLS estimates of SSP effects on participants are such that effects of a plausible magnitude would likely come out as significant. Of course, smaller effects, say 0.15σ, would not.

Finally, the over-identification test statistics in Table 8 reject the hypothesis of SFP and SFSP treatment-effect equality (or come close). Substantively, this means that the differences in intention-to-treat estimates reported in earlier tables are at least marginally significant after adusting for differences in compliance rates.

IV. Student Reports

In order to better understand students' perceptions of the STAR program and their reaction to it, we conducted open-ended interviews with participants in each of the treatment groups. Interviewees were chosen randomly and offered $20 university bookstore gift certificates for attending. We contacted 54 students by e-mail and phone to obtain a total of 10 interviewees, 7 of whom were female. The students were interviewed focus-group style, separately for each treatment group (two from SFSP, five from SFSP, and three from SFSP), and could react to statements by others in the group. Interviews lasted about one hour and were guided by a list of questions. We also conducted one-on-one interviews lasting 30 minutes with 5 women and 1 man. (We invited 11 SFP and SFSP participants for one-on-one interviews and 6 responded).

Most students described their initial reaction to the fellowship program as strongly positive, though a few wondered at first if the program "was too good to be true." One participant commented, "A couple people I mentioned it to described it as a bribe to do well, as motivation, but hey, it worked for me." Another commented, "I wanted it. $5,000! I definitely wasn't going to miss that." (This student received the full fellowship.) A male student we spoke with, however, said the money was not that important a motivator because "I already have enough for 1st and 2nd year tuition so finding the money isn't a big problem for me."

Sentiment about the fellowship program as the year progressed was mixed. Some felt the program helped motivate them to improve. One student reported that "I found my study habits improved. It might have been the money motivation." Another student said, "First semester was a rude awakening for me because the work is very different than high school…Project STAR definitely gave me the focus and motivation." [22] Another felt that "It helped a lot. That promise of money really helped to motivate me. I felt like I was being paid to go to school. It helped me to jump in…think about all my classes and do well in all of them rather than letting myself slip and focus more on ones that I enjoy and leaving the others in the dust." Other students, however, became more concerned about poor grades than good grades and generally forgot about the fellowship program. One SFSP student who finished with a 1.2 GPA commented, "At first I was excited about it [the program], but when I was in school I kind of forgot….The [fellowship] I think was good, but I didn't really focus on it. I was more worried about my grades." Another student commented, "I thought about it [the SFP] a lot in first semester. But then when I realized my grades weren't going anywhere, I just forgot about it. But in first semester, yeah, I thought about it all the time." Interestingly, no one said the program detracted from their first year experience.

Some interviewees from the fellowship groups suggested that the fellowship reminders through e-mails and advisor contacts mattered quite a bit. No student said there were too many fellowship reminders or too many e-mails from advisors.

Those we talked to about the SSP focused almost exclusively on the advisor program. Many students were pleased with their advisor interactions, or were simply glad to know that they could communicate with their advisor if needed. One male SSP students noted, "University is something different and I also wanted extra help.

[22] Asked why this student found university so different from high school, she responded, "I could do things last minute (in high school). I always did everything the night before and I got straight A's. So to come here and then to basically fail every subject was like, 'oh my gosh, like what am I doing?' It's crazy—it's extremely stressful—it's unbelievable."

The peer advisor, I personally did not meet him, but he was really helpful because at every 15 days he used to e-mail me and ask me how it's going and how I did on the test." Another female student said, "I thought that it was nice to know that there was someone there that you could run to and ask for help. At the beginning I just started to talk to my advisor and she did give me some advice but I found that at the time it's hard to just all of a sudden change all of your schedules just because she said you should have good time management and you should do this and that. So I guess that you slowly start to see what your peer advisor is talking about then you try to change for the best." Another student never met or responded to her advisor, but nevertheless felt the advisor's regular e-mails were helpful, "Like somebody who cared."

These discussions suggest that the fellowship program penetrated students' consciousness, but in some cases interest declined over the course of the school year. Reminders and communication from advisors helped maintain interest and motivation. Surprisingly, the students we spoke with did not seem to view the fellowship targets as unrealistic. Some who did not receive a fellowship still felt that they benefited from the fellowship program, and no one felt the program influenced them in a negative way. On the service side, virtually all participants welcomed interaction with upper-year student advisors, even when only through e-mail.

V. Summary and Conclusions

The goal of the Student Achievement and Retention (STAR) Project was to learn more about the potential for support services and financial incentives to improve academic performance in college. Student interest in support services was lower than expected. On the other hand, interest in services as reflected in sign-up rates and service usage was markedly higher in the group that was also offered cash incentives. Interest in services and use of services was also much higher for young women than young men. Peer advising was considerably more popular than supplemental instruction for both sexes. The peer-advising intervention clearly bears further exploration, as does the use of achievement incentives to boost interest in services.

A number of patterns emerge from the STAR results. First, students offered services without fellowships did no better than those in the control group. This may be because sign-up rates were relatively low in the treatment groups offered services, since low consent rates dilute intention-to-treat effects. On the other hand, a 2SLS analysis that adjusts intention-to-treat effects for nonparticipation reveals a level of precision sufficient to detect theoretical service effects equal to about 0.25σ in the combined sample of men and women.

Although we observed an initial boost to women's fall grades in both the fellowship only and combined groups, the only significant achievement gains by the end of the first year were for women offered both fellowships and services through the SFSP. Also noteworthy is that even though STAR incentives and services were available for only one year, the significant treatment effect observed for SFSP females persisted into the second year. The average GPA for SFSP females in the second year was about 0.28 higher than the GPA in the control group, a substantial and statistically significant effect. SFSP females also earned a quarter credit more than controls in the second year, so that after two years, the SFSP-control difference amounted to a little over half a credit. This is important because it shows that incentives, in combination with increased services (especially, it would seem, peer advising), changed behavior in a manner consistent with a lasting improvement in study skills.

The costs incurred for these long-lasting, SFSP-induced achievement gains were fairly modest. Total STAR expenditure on services was about \$121,000, or about \$302 for each of 400 treated participants in the SSP and SFSP.[23] Total expenditure on fellowships was \$146,500 or \$366 for each of 400 treated participants in the SFP and SFSP. Individual fellowship costs were higher in the SFSP than the SFP because the award rate was higher in the SFSP group. The total SFSP fellowship cost was \$65,500, spread over 150 participants, or about \$437 per participant. The average cost per SFSP participant was \$739, low by the standards of other social programs.[24] In future work, we plan to look at lower-cost service strategies that focus more on e-mail-based peer advising. Even at the current cost level, however, it seems likely that for women these amounts will be more than offset by future earnings gains should the effects reported here translate into increased completion rates and better academic performance down the road. For men, on the other hand, the program was clearly ineffective.

This raises the important question of why women responded so much more to the SFSP than men. Although we have no simple explanation for this difference, it is worth noting that women now outperform men on many measures of academic success. Women generally receive higher grades beginning in late primary school, have fewer disciplinary problems, are less likely to repeat grades, and are less likely to report that they don't like school (Jusith Kleinfeld 1998). Clauda Goldin, Lawrence F. Katz, and Ilyana Kuziemko (2006) and Marc Frenette and Klarka Zeman (2007) note that women have substantially higher college attendance and completion rates than men in the United States and Canada. Jacob (2002) attributes many sex differences in college attendance and completion to differences in noncognitive ability and suggests women may have more foresight when making decisions with long-term consequences. Specifically, women appear to have better study habits, may be more motivated to do well in school, and therefore to take advantage of programs like STAR.

Finally, it bears repeating that similar sex differentials in program impact have been observed elsewhere. Dynarski (2005) estimates larger effects of tuition aid on college completion for women in the United States, while Garibaldi et al. (2007) find that tuition affects the completion rates of women more than men in Italy. In a study of the effects of merit awards on Israeli high school students, Angrist and Lavy (2002) find effects on girls only. A more modest but still marked gender differential crops up in the response to randomly assigned vouchers for private secondary schools in Colombia (Angrist et al. 2002). Farther afield, Michael Anderson's (2006) evaluation of three preschool programs suggests these programs benefit girls but

[23] This is the sum of \$45,292 in pay and benefits for advisors, \$57,573 in pay and benefits for facilitators, and \$17,789 for supplies and operating costs.

[24] For example, short-term training programs such as the JTPA and Job Corps typically spend thousands of dollars per participant (see, e.g., Robert J. Lalonde 1995).

not boys, and the Moving to Opportunity (MTO) evaluation (e.g., Susan Clampet-Lundquist et al. 2006) points to benefits of subsidized housing in nonpoverty areas for women but negative effects on men. These gender differences in the response to incentives and services constitute an important area for further study.

Appendix

Student Fellowship Program Award Schedule

	Award amount		
High school GPA quartile	$1,000	$2,500	$5,000
0–25th percentile	2.3 (C+)	2.7 (B−)	3.0 (B)
25th–50th percentile	2.7 (B−)	3.0 (B)	3.3 (B+)
50th–75th percentile	3.0 (B)	3.3 (B+)	3.7 (A−)

Notes: Eligibillity was determined by the student's best four courses. Half of SFP/SFSP participants were offered the opportunity to qualify for a $2,500 award.

REFERENCES

Anderson, Michael. 2006. "Uncovering Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." Ph.D. Diss. MIT Department of Economics.

Angrist, Joshua D. 1991. "Grouped-Data Estimation and Testing in Simple Labor-Supply Models." *Journal of Econometrics,* 47(2–3): 243–66.

Angrist, Joshua, Daniel Lang, Leonard Leung, Philip Oreopoulos, and Yan Tam-Seguin. 2008. "The Student Achievement and Retention Project: An Experiment to Evaluate the Impacts of Offering Additional Services and Merit-Scholarships to First-Year Undergraduates in Canada." Report submitted to Canada Millennium Scholarship Foundation.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2006. "Lead Them to Water and Pay Them to Drink: An Experiment with Services and Incentives for College Achievement." National Bureau of Economic Research Working Paper 12790.

Angrist, Joshua D., Daniel Lang, and, Philip Oreopoulos. 2007. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." IZA Discussion Paper 3134.

Angrist, Joshua D., and Victor Lavy. 2002. "The Effect of High School Matriculation Awards: Evidence from Randomized Trials." National Bureau of Economic Research Working Paper 9389.

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

Arendale, David R. 2001. "Supplemental Instruction (SI): Review of Research Concerning the Effectiveness of SI from The University of Missouri-Kansas City and Other Institutions from Across the United States." Unpublished.

Ashworth, Karl, Jay Hardman, Woon-Chia Liu, Sue Maguire, Sue Middleton, Lorraine Dearden, Carl Emmerson, and Christine Frayne. 2001. "Education Maintenance Allowance: The First Year, A Quantitative Evaluation." UK Department for Education and Evaluation Research, Brief 257, May.

Bailey, Thomas R., and Mariana Alfonso. 2005. "Paths to Persistence: An Analysis of Research on Program Effectiveness at Community Colleges." *Lumina Foundation for Education, New Agenda Series,* 6(1). http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/a1/fd.pdf

Barefoot, Betsey O. 2004. "Higher Education's Revolving Door: Confronting the Problem of Student Drop-out in US Colleges and Universities." *Open Learning,* 19(1): 9–18.

Bettinger, Eric P., and Bridget Terry Long. 2005. "Addressing the Needs of Under-Prepared Students in Higher Education: Does College Remediation Work?" National Bureau of Economic Research Working Paper 11325.

Bloom, Dan, and Colleen Sommo. 2005. "Building Learning Communities: Early Results from the Opening Doors Demonstration at Kingsborough Community College." MDRC Report, June. http://www.mdrc.org/publications/410/full.pdf.

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

Brock, Thomas, and Lashawn Richburg-Hayes. 2006. "Paying for Persistence: Early Results of a Louisiana Scholarship Program for Low-Income Parents Attending Community College." MDRC Report, May. http://www.mdrc.org/publications/429/full.pdf.

Card, David. 1995. "Earnings, Schooling, and Ability Revisited," In *Research in Labor Economics*, Vol. 14., ed. Solomon W. Polachek, 23–48. Greenwich, CT: JAI Press.

Charness, Gary, and Uri Gneezy. 2006. "Incentives and Habits." Unpublished.

Clampet-Lundquist, Susan, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2006. "Moving At-Risk Teenagers Out of High-Risk Neighborhoods: Why Girls Fare Better Than Boys." Princeton Industrial Relation Section Working Paper 509.

Congos, Dennis H., and N. Schoeps. 2003. "Inside Supplemental Instruction (SI): One Model of What Happens that Improves Grades and Retention Revisited." *Journal of Student Centered Learning,* 1(13): 159–70.

Consortium for Student Retention Data Exchange. 2004. "2003-04 CSRDE Report: The Retention and Raduation Rates in 344 Colleges and Universities." Center for Institutional Data Exchange and Analysis, University of Oklahoma Outreach.

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.

DesJardins, Stephen L., and Brian P. McCall. 2006. "The Impact of the Gates Millennium Scholarship Program on the Retention and Academic Performance of Minority Students: A Regression-Discontinuity Analysis." Unpublished.

Dominitz, Jeff, and Charles F. Manski. 1997. "Using Expectations Data to Study Subjective Income Expectations." *Journal of the American Statistical Association*, 92(439): 855–67.

Dynarski, Susan. 2002. "The Consequences of Merit Aid." National Bureau of Economic Research Working Paper 9400.

Dynarski, Susan. 2005. "Building the Stock of College-Educated Labor." National Bureau of Economic Research Working Paper 11604.

Firpo, Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica*, 75(1): 259–76.

Frederick, Shane, George Loewenstein, and Ted O' Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 40(2): 351–401.

Frenette, Marc, and Klarka Zeman. 2007. "Why are Most University Students Women? Evidence Based on Academic Performance, Study Habits, and Parental Influences." Statistics Canada Analytical Studies Branch Research Paper 303.

Frölich, Markus, and Blaise Melly. 2008. "Unconditional Quantile Treatment Effects Under Endogeneity." IZA Discussion Paper 3288.

Fudenberg, Drew, and David K. Levine. 2006. "A Dual-Self Model of Impulse Control." *American Economic Review*, 96(5): 1449–76.

Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Rettore. 2007. "College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities." National Bureau of Economic Research Working Paper 12863.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives*, 20(4): 133–56.

Goodlad, John I. 2004. *A Place Called School: Prospects for the Future*. 20th Anniversary ed. New York: McGraw-Hill.

Grossman, Jean Baldwin, and Joseph P. Tierney. 1998. "Does Mentoring Work? An Impact Study of the Big Brothers Big Sisters Program." *Evaluation Review,* 22(3): 403–26.

Habley, Wesley R., ed. 2004. *The Status of Academic Advising: Findings from the ACT Sixth National Survey* (Monograph No. 10). Manhattan, KS: National Academic Advising Association.

Hensen, Kari A., and Mack Shelley. 2003. "The Impact of Supplemental Instruction: Results from a large, public, Midwestern University." *Journal of College Student Development,* 44(2): 250–59.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–75.

Jackson, C. Kirabo. 2007. "A Little Now for a Lot Later: A Look at the Texas Advanced Placement Incentive Program." Cornell University, Cornell Higher Education Research Institute Working Paper 107.

Jacob, Brian A. 2002. "Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review*, 21(6): 589–98.

Kane, Thomas J., and Cecilia Elena Rouse. 1995. "Labor-Market Returns to Two- and Four-Year College." *American Economic Review*, 85(3): 600–614.

Kleinfeld, Judith. 1998. "The Myth That Schools Shortchange Girls: Social Science in the Service of Deception." Washington, DC: Women's Freedom Network.

Kremer Michael, Edward Miguel, and Rebecca Thornton. 2004. "Incentives to Learn." Center for International Development Working Paper 109.

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

Lalonde, Robert J. 1995. "The Promise of U.S. Employment and Training Programs." *Journal of Economic Perspectives*, 9(2): 149–68.

Lavy, Victor, and Analia Schlosser. 2005. "Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits." *Journal of Labor Economics*, 23(4): 839–74.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. 2003. "The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment." CEPR Discussion Paper 3921.

Lotkowski, Veronica A., Steven B. Robbins, and Richard J. Noeth. 2004. "The Role of Academic and Non-academic Factors in Improving College Retention." American College Testing (ACT) Policy Report. http://www.act.org/research/policymakers/pdf/college_retention.pdf.

Medina, Jennifer. 2007a. "Schools Plan to Pay Cash for Marks," *The New York Times,* June 19.

Medina, Jennifer. 2007b. "Making Cash a Prize for High Scores on Advanced Placement Tests," *The New York Times,* October 15.

National Center for Education Statistics. 2003. "Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000." Washington DC: Department of Education. http://nces.ed.gov/pubs2004/2004010.pdf.

O' Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review*, 89(1): 103–24.

Ogden, Peggy, Dennis Thompson, H. A. Russell and Carol Simons. 2003. "The Short- and Long-term Impact of Supplemental Instruction." *Journal of Developmental Education,* 26(3): 2–9.

Oreopoulos, Philip. 2007. "Do Dropouts Drop out Too Soon? Wealth, Health and Happiness from Compulsory Schooling." *Journal of Public Economics*, 91(11–12): 2213–29.

Pascarella, Ernest T., and Patrick T. Terenzini. 1991. *How College Affects Students: Findings and Insights from Twenty Years of Research*. San Francisco: Jossey-Bass.

Seligman, Jason, Richard Milford, John O'Looney, and Jim Ledbetter. 2004. "HOPE Scholarship Joint Study Commission Report." Athens, GA: Carl Vinson Institute of Government.

Spear, L. P. 2000. "The Adolescent Brain and Age-Related Behavioral Manifestations," *Neuroscience and Biobehavioral Reviews,* 24(4): 417–63.

Tinto, Vincent. 1993. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. 2nd ed. Chicago: University of Chicago Press.

Wyckoff, Susan C. 1998. "Retention Theories in Higher Education: Implications for Institutional Practice." *Recruitment and Retention in Higher Education,* 12(2): 2–7.

This article has been cited by:

1. Oded Gurantz. 2018. A Little Can Go a Long Way: The Impact of Advertising Services on Program Take-Up. *Educational Evaluation and Policy Analysis* **40**:3, 382-398. [Crossref]
2. William H. Yeaton, Brian G. Moss. 2018. A Multiple-Design, Experimental Strategy: Academic Probation Warning Letter's Impact on Student Achievement. *The Journal of Experimental Education* 1-21. [Crossref]
3. Eline Sneyers, Kristof De Witte. 2018. Interventions in higher education and their effect on student success: a meta-analysis. *Educational Review* **70**:2, 208-228. [Crossref]
4. Dennis A. Kramer, Michael R. Holcomb, Robert Kelchen. 2018. The Costs and Consequences of Excess Credit Hours Policies. *Educational Evaluation and Policy Analysis* **40**:1, 3-28. [Crossref]
5. Lester Lusher, Doug Campbell, Scott Carrell. 2018. TAs like me: Racial interactions between graduate teaching assistants and undergraduates. *Journal of Public Economics* **159**, 203-224. [Crossref]
6. Arnaud Chevalier, Peter Dolton, Melanie Lührmann. 2018. 'Making it count': incentives, student effort and performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **181**:2, 323-349. [Crossref]
7. Jenna Howard Terrell, Daniel Bugler. 2017. If you build it, will they come? Lessons learned in recruiting students for randomized controlled trials in postsecondary settings. *Educational Research and Evaluation* **23**:7-8, 272-289. [Crossref]
8. Fenella Carpena, Shawn Cole, Jeremy Shapiro, Bilal Zia. 2017. The ABCs of Financial Education: Experimental Evidence on Attitudes, Behavior, and Cognitive Biases. *Management Science* . [Crossref]
9. Leonardo Bursztyn, Robert Jensen. 2017. Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure. *Annual Review of Economics* **9**:1, 131-153. [Crossref]
10. Lisa Barrow, Cecilia Elena Rouse. 2017. Financial Incentives and Educational Investment: The Impact of Performance-based Scholarships on Student Time Use. *Education Finance and Policy* 1-30. [Crossref]
11. Brent J. Evans. 2017. SMART Money: Do Financial Incentives Encourage College Students to Study Science?. *Education Finance and Policy* **12**:3, 342-368. [Crossref]
12. Scott Carrell, Bruce Sacerdote. 2017. Why Do College-Going Interventions Work?. *American Economic Journal: Applied Economics* **9**:3, 124-151. [Abstract] [View PDF article] [PDF with links]
13. Andreas Ostermaier. 2017. Incentives for students: effects of certificates and deadlines on student performance. *Journal of Business Economics* **96**. . [Crossref]
14. Malte Sandner, Tanja Jungmann. 2017. Gender-specific effects of early childhood intervention: Evidence from a randomized controlled trial. *Labour Economics* **45**, 59-78. [Crossref]
15. Valentina Paredes. 2017. Grading System and Student Effort. *Education Finance and Policy* **12**:1, 107-128. [Crossref]
16. R.G. Fryer. The Production of Human Capital in Developed Countries 95-322. [Crossref]
17. D. Xu. 2016. Assistance or Obstacle? The Impact of Different Levels of English Developmental Education on Underprepared Students in Community Colleges. *Educational Researcher* . [Crossref]

18. Steven D. Levitt, John A. List, Susanne Neckermann, Sally Sadoff. 2016. The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. *American Economic Journal: Economic Policy* **8**:4, 183-219. [Abstract] [View PDF article] [PDF with links]

19. Eva Argente-Linares, M. Carmen Pérez-López, Celia Ordóñez-Solana. 2016. Practical Experience of Blended Mentoring in Higher Education. *Mentoring & Tutoring: Partnership in Learning* **24**:5, 399-414. [Crossref]

20. Danielle R. Adams, Steven A. Meyers, Rinad S. Beidas. 2016. The relationship between financial strain, perceived stress, psychological symptoms, and academic and social integration in undergraduate students. *Journal of American College Health* **64**:5, 362-370. [Crossref]

21. Alexander K. Mayer, Reshma Patel, Melvin Gutierrez. 2016. Four-Year Degree and Employment Findings From a Randomized Controlled Trial of a One-Year Performance-Based Scholarship Program in Ohio. *Journal of Research on Educational Effectiveness* **9**:3, 283-306. [Crossref]

22. Sara Goldrick-Rab, Robert Kelchen, Douglas N. Harris, James Benson. 2016. Reducing Income Inequality in Educational Attainment: Experimental Evidence on the Impact of Financial Aid on College Completion. *American Journal of Sociology* **121**:6, 1762-1817. [Crossref]

23. Joshua D. Angrist, Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, Christopher R. Walters. 2016. Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice. *Journal of Labor Economics* **34**:2, 275-318. [Crossref]

24. A.M. Lavecchia, H. Liu, P. Oreopoulos. Behavioral Economics of Education 1-74. [Crossref]

25. Cynthia Miller, James Riccio, Nandita Verma, Stephen Nuñez, Nadine Dechausay, Edith Yang. 2015. Testing a conditional cash transfer program in the U.S.: the effects of the family rewards program in New York City. *IZA Journal of Labor Policy* **4**:1. . [Crossref]

26. Miguel Urquiola. 2015. Progress and challenges in achieving an evidence-based education policy in Latin America and the Caribbean. *Latin American Economic Review* **24**:1. . [Crossref]

27. JESSICA VECHBANYONGRATANA, SASIWIMON WARUNSIRI PAWEENAWAT. 2015. TRANSFER PAYMENTS AND UPPER SECONDARY SCHOOL OUTCOMES: THE CASE OF LOW-INCOME FEMALE STUDENTS IN THAILAND. *The Singapore Economic Review* **60**:05, 1550082. [Crossref]

28. Matthew G. Springer, Brooks A. Rosenquist, Walker A. Swain. 2015. Monetary and Nonmonetary Student Incentives for Tutoring Services: A Randomized Controlled Trial. *Journal of Research on Educational Effectiveness* **8**:4, 453-474. [Crossref]

29. Matthew C. Rousu, Jay R. Corrigan, David Harris, Jill K. Hayter, Scott Houser, Becky A. Lafrancois, Olugbenga Onafowora, Gregory Colson, Adam Hoffer. 2015. Do Monetary Incentives Matter in Classroom Experiments? Effects on Course Performance. *The Journal of Economic Education* **46**:4, 341-349. [Crossref]

30. Brian G. Moss, William H. Yeaton. 2015. Failed Warnings. *Evaluation Review* **39**:5, 501-524. [Crossref]

31. Maria De Paola, Vincenzo Scoppa. 2015. Procrastination, academic success and the effectiveness of a remedial program. *Journal of Economic Behavior & Organization* **115**, 217-236. [Crossref]

32. J Cody Davidson. 2015. The Effects of a State Need-based Access Grant on Traditional and Non-traditional Student Persistence. *Higher Education Policy* **28**:2, 235-257. [Crossref]

33. Oriana Bandiera, Valentino Larcinese, Imran Rasul. 2015. Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics* **34**, 13-25. [Crossref]

34. Gabrielle Fack, Julien Grenet. 2015. Improving College Access and Success for Low-Income Students: Evidence from a Large Need-Based Grant Program. *American Economic Journal: Applied Economics* **7**:2, 1-34. [Abstract] [View PDF article] [PDF with links]
35. Louis-Philippe Morin. 2015. Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform. *Journal of Labor Economics* **33**:2, 443-491. [Crossref]
36. Hans Bonesrønning, Leiv Opstad. 2015. Can student effort be manipulated? Does it matter?. *Applied Economics* **47**:15, 1511-1524. [Crossref]
37. Karin Barac. 2015. Helping Disadvantaged Students: Findings from the Thuthuka Programme. *Accounting Education* **24**:2, 75-101. [Crossref]
38. Timothy M Diette, Manu Raghav. 2015. Class Size Matters: Heterogeneous Effects of Larger Classes on College Student Learning. *Eastern Economic Journal* **41**:2, 273-283. [Crossref]
39. Malte Sandner. 2015. The effects of high-quality student mentoring. *Economics Letters* . [Crossref]
40. Rashmi Barua, Marian Vidal-Fernandez. 2014. No Pass No Drive: Education and Allocation of Time. *Journal of Human Capital* **8**:4, 399-431. [Crossref]
41. Markus Frölich, Martin Huber. 2014. Treatment evaluation with multiple outcome periods under endogeneity and attrition. *Journal of the American Statistical Association* 00-00. [Crossref]
42. Martin Huber. 2014. Treatment Evaluation in the Presence of Sample Selection. *Econometric Reviews* **33**:8, 869-905. [Crossref]
43. Moussa P. Blimpo. 2014. Team Incentives for Education in Developing Countries: A Randomized Field Experiment in Benin. *American Economic Journal: Applied Economics* **6**:4, 90-109. [Abstract] [View PDF article] [PDF with links]
44. Yvonne Oswald, Uschi Backes-Gellner. 2014. Learning for a bonus: How financial incentives interact with preferences. *Journal of Public Economics* **118**, 52-61. [Crossref]
45. Thomas Bolli, Geraint Johnes. 2014. In my own time: tuition fees, class time and student effort in non-formal (or continuing) education. *Journal of Education and Work* 1-15. [Crossref]
46. Martin Huber, Giovanni Mellace. 2014. Testing exclusion restrictions and additive separability in sample selection models. *Empirical Economics* **47**:1, 75-92. [Crossref]
47. Lisa Barrow, Lashawn Richburg-Hayes, Cecilia Elena Rouse, 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. [Crossref]
48. Kevin Denny, Orla Doyle, Patricia McMullin, Vincent O'Sullivan. 2014. Money, mentoring and making friends: The impact of a multidimensional access program on student performance. *Economics of Education Review* **40**, 167-182. [Crossref]
49. Ashwin Satyanarayana, Hong Li, Josephine Braneky. Improving retention by mentoring and tutoring freshmen students 1-5. [Crossref]
50. David J. Deming, Justine S. Hastings, Thomas J. Kane, Douglas O. Staiger. 2014. School Choice, School Quality, and Postsecondary Attainment. *American Economic Review* **104**:3, 991-1013. [Abstract] [View PDF article] [PDF with links]
51. Tao Li, Li Han, Linxiu Zhang, Scott Rozelle. 2014. Encouraging classroom peer interactions: Evidence from Chinese migrant schools. *Journal of Public Economics* **111**, 29-45. [Crossref]
52. Eric P. Bettinger, Rachel B. Baker. 2014. The Effects of Student Coaching. *Educational Evaluation and Policy Analysis* **36**:1, 3-19. [Crossref]

53. Rajeev Darolia. 2014. Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review* **38**, 38-50. [Crossref]

54. S. Sadoff. 2014. The role of experimentation in education policy. *Oxford Review of Economic Policy* **30**:4, 597. [Crossref]

55. C. KIRABO JACKSON. 2013. DO COLLEGE-PREPARATORY PROGRAMS IMPROVE LONG-TERM OUTCOMES?. *Economic Inquiry* no-no. [Crossref]

56. Stephen Gibbons, Olmo Silva, Felix Weinhardt. 2013. Everybody Needs Good Neighbours? Evidence from Students' Outcomes in England. *The Economic Journal* **123**:571, 831-874. [Crossref]

57. K. F. Butcher, M. G. Visher. 2013. The Impact of a Classroom-Based Guidance Program on Student Performance in Community College Math Classes. *Educational Evaluation and Policy Analysis* **35**:3, 298-323. [Crossref]

58. Lex Borghans, Huub Meijers, Bas ter Weel. 2013. The importance of intrinsic and extrinsic motivation for measuring IQ. *Economics of Education Review* **34**, 17-28. [Crossref]

59. Victor Lavy. 2013. Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments among Teachers*. *The Economic Journal* **123**:569, 540-573. [Crossref]

60. Louis-Philippe Morin. 2013. Estimating the benefit of high school for university-bound students: evidence of subject-specific human capital accumulation. *Canadian Journal of Economics/Revue canadienne d'économique* **46**:2, 441-468. [Crossref]

61. Núria Rodríguez-Planas. 2012. Mentoring, educational services, and incentives to learn: What do we know about them?. *Evaluation and Program Planning* **35**:4, 481-490. [Crossref]

62. Jason M. Lindo,, Isaac D. Swensen,, Glen R. Waddell. 2012. Are Big-Time Sports a Threat to Student Achievement?. *American Economic Journal: Applied Economics* **4**:4, 254-274. [Abstract] [View PDF article] [PDF with links]

63. Núria Rodríguez-Planas. 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, 121-139. [Abstract] [View PDF article] [PDF with links]

64. Pietro Garibaldi, Francesco Giavazzi, Andrea Ichino, Enrico Rettore. 2012. College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities. *Review of Economics and Statistics* **94**:3, 699-711. [Crossref]

65. Douglas N. Harris, Sara Goldrick-Rab. 2012. Improving the Productivity of Education Experiments: Lessons from a Randomized Study of Need-Based Financial Aid. *Education Finance and Policy* **7**:2, 143-169. [Crossref]

66. Victor Lavy, Olmo Silva, Felix Weinhardt. 2012. The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics* **30**:2, 367-414. [Crossref]

67. Maria De Paola, Vincenzo Scoppa, Rosanna Nisticó. 2012. Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment. *Journal of Human Capital* **6**:1, 56-85. [Crossref]

68. Marco Castillo, Paul J. Ferraro, Jeffrey L. Jordan, Ragan Petrie. 2011. The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics* **95**:11-12, 1377-1385. [Crossref]

69. R. G. Fryer. 2011. Financial Incentives and Student Achievement: Evidence From Randomized Trials. *The Quarterly Journal of Economics* . [Crossref]

70. Kevin Milligan,, Mark Stabile. 2011. Do Child Tax Benefits Affect the Well-being of Children? Evidence from Canadian Child Benefit Expansions. *American Economic Journal: Economic Policy* **3**:3, 175-205. [Abstract] [View PDF article] [PDF with links]
71. Thomas S. Dee. 2011. Conditional cash penalties in education: Evidence from the Learnfare experiment. *Economics of Education Review* . [Crossref]
72. Nicholas Burger, Gary Charness, John Lynham. 2011. Field and online experiments on self-control. *Journal of Economic Behavior & Organization* **77**:3, 393-404. [Crossref]
73. John Bound, Sarah Turner. Dropouts and Diplomas 573-613. [Crossref]
74. C. Kirabo Jackson. 2010. Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago*. *The Economic Journal* **120**:549, 1399-1429. [Crossref]
75. Maria De Paola, Vincenzo Scoppa. 2010. A signalling model of school grades under different evaluation systems. *Journal of Economics* **101**:3, 199-212. [Crossref]
76. Gary S. Becker, William H. J. Hubbard, Kevin M. Murphy. 2010. Explaining the Worldwide Boom in Higher Education of Women. *Journal of Human Capital* **4**:3, 203-241. [Crossref]
77. Paco Martorell, Isaac McFarlin. 2010. Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes. *Review of Economics and Statistics* 110223123706012. [Crossref]
78. Jason M. Lindo,, Nicholas J. Sanders,, Philip Oreopoulos. 2010. Ability, Gender, and Performance Standards: Evidence from Academic Probation. *American Economic Journal: Applied Economics* **2**:2, 95-117. [Abstract] [View PDF article] [PDF with links]
79. Rey Hernández-Julián. 2010. Merit-Based Scholarships and Student Effort. *Education Finance and Policy* **5**:1, 14-35. [Crossref]
80. Abhijit V. Banerjee, Esther Duflo. 2009. The Experimental Approach to Development Economics. *Annual Review of Economics* **1**:1, 151-178. [Crossref]
81. Joshua Angrist,, Victor Lavy. 2009. The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial. *American Economic Review* **99**:4, 1384-1414. [Abstract] [View PDF article] [PDF with links]