# MIT Open Access Articles

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

INCENTIVES AND SERVICES FOR COLLEGE ACHIEVEMENT:

EVIDENCE FROM A RANDOMIZED TRIAL

By,

Joshua Angrist
MIT, NBER

Daniel Lang
University of Toronto

Philip Oreopoulos
University of Toronto, NBER

May 2008

INCENTIVES AND SERVICES FOR COLLEGE ACHIEVEMENT:

EVIDENCE FROM A RANDOMIZED TRIAL

Abstract

Many North American college students have trouble satisfying degree requirements in a timely manner.  This paper reports on a randomized field experiment involving two strategies designed to improve academic performance among entering full-time undergraduates at a large Canadian university. One treatment group ("services") was offered peer advising and organized study groups.  Another ("incentives") was offered substantial merit-scholarships for solid, but not necessarily top, first year grades.  A third treatment group combined both interventions, while a control group received neither services nor incentives.  Service take-up rates were much higher for women than for men and for students offered both services and incentives than for those offered services alone.  No program had an effect on men's grades or other measures of academic performance.  However, the Fall and first-year grades of women in the combined group were higher than those of women in the control group, and women in this group earned more course credits and were less likely than controls to be on academic probation.  These differentials persisted through the end of the second year, in spite of the fact that incentives were given in the first year only.  The results suggest that the study skills acquired in response to a combination of academic support services and incentives can have a lasting effect, at least on women, and that the combination of services and incentives is more promising than either alone.

Joshua Angrist
Department of Economics
MIT
angrist@mit.edu

Daniel Lang
Ontario Institute for Studies in Education
The University of Toronto
dlang@oise.utoronto.ca

Philip Oreopoulos
Department of Economics
The University of Toronto
oreo@economics.utoronto.ca

I. Introduction

Recent years have seen growing interest in interventions designed to increase college attendance and completion, especially for low-income students. Major efforts to increase enrolment 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., Kane and Rouse, 1995).

In addition to the obvious necessity of starting college, an important part of the post-secondary education process is academic performance. Many students struggle and take much longer to attain a degree than the nominal completion time. First-year students are especially likely to run into trouble. Nearly one-third of first-year college students in the U.S. 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 (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 (Tinto, 1993; Goodlad, 2004).

Like academic support services, merit-scholarships have a long history in the post-secondary context, but traditional programs, such as the U.S. 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 include state tuition waivers for students who maintain a B-average, such as Georgia's HOPE Program. As Dynarski (2005) notes, programs of this sort are relevant for many students. For example, nearly 60% 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 (Seligman et al. 2004). The promise of a scholarship may increase the time devoted to schoolwork and lead students to acquire better study habits.

Non-experimental evidence 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 Bailey and Alfonso, 2005, Pascarella and Terenzini, 1991, Lotkowski, Robbins, and Noeth, 2004, and 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., Grossman and Tierney (1998), Lavy and Schlosser (2005), Bettinger and Long (2005), and Bloom and 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; Cornwell, Mustard, and Sridhar, 2006). A few recent studies look at the impact of financial incentives on

---

[1]The National Merit program awards roughly 8200 scholarships to students selected from 1.4 million PSAT-takers.

the performance of college students, though some of this research is still on-going (e.g. DesJardins and McCall (2006), and Brock and Richburg-Hayes (2006)).[2] An evaluation of gradually increasing tuition payments in response to delayed completion by Garibaldi, et al. (2007) found substantial effects on Italian college women, while Leuven, Oosterbeek and 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, with 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 using any sort of research design. The control group was eligible for standard university support services but received nothing extra.

The first finding that comes out of Project STAR is much higher use of services by young women than young men. This is consistent with a range of evidence pointing to smoother

---

[2] Other evidence on incentives for academic performance comes from pre-college populations. For example, Ashworth et al. (2001) discuss a non-experimental evaluation of stipends for high school students who stay in school and Kremer, Miguel, and 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. Finally, Angrist et al. (2002) evaluate the impact of school vouchers in Colombia that required students meet grade promotion standards for eligibility. All of these programs point to at least some positive effects for some types of primary or secondary school students.

college careers for women (see, e.g., 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, particular in one treatment group.

The effects on women are strongest among 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, the initial grade boost faded for the fellowship-only group so that 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 the rest of the STAR population in second year. This is in spite of the fact that fellowships and services were available in first year only. These results suggest 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 undergraduate students 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 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 successfully with the first year of school. Advisors emailed participants regularly and were available to meet at the STAR office. FSGs are class-specific sessions designed to improve students' study habits and learning strategies, 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 (Arendale, 2001).

The SFP offered 250 students the opportunity to win merit scholarships for maintaining solid but not necessarily top grades in first year. Award targets were set based on high school grade quartiles. Participants from the lowest grade quartile were offered $5,000 in cash, roughly a year's tuition, for attaining a grade average of B (a GPA of 3.0) or higher, or $1,000 in cash for a C+ (a GPA of 2.3).[3] Participants from the second and third quartiles were offered $5,000 awards for attaining a grade average of B+ or A- respectively, or $1,000 awards for attaining a B- or B average respectively. Upper-quartile students were not eligible to participate in the study. To qualify for a fellowship, SFP students had to take at least 4 courses per term and

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

register to attend the second year of their program (a full load, required to complete a degree program in four years, is 5 courses per year). In the 2003-4 school year, 7-8 percent of registered students met the standard for a $5,000 award, while 26-28 percent met the standard for a $1,000 award. As it turns out, however, award rates in our cohort were somewhat lower.

A third treated group of 150 students was offered both the SSP and SFP. It is important to note, however, that other than being given 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 1006 students, with whom program operators had no contact beyond a baseline survey that went to all incoming freshman.[4]

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 (Habley, 2004). Several universities match first year students with upper-class peers or faculty advisors who provide academic support. 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 3-day training course and received continuous training and feedback from supervisors. Advisors emailed 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 informing advisees about 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

---

[4]16 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 non-program students.

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 are voluntary, course-focused, weekly sessions open to all treated students. FSG facilitators are students who were previously successful in the course they were hired to facilitate. They attend the course with their assigned STAR students, and try to help students develop reasoning skills useful for the subject they are facilitating. FSGs are 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, Noeth, 2004). A number of studies suggest students who participate in FSG-style supplemental instruction outperform non-participating peers (Congos and Schoeps, 2003, Hensen and Shelley, 2003, 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 can get 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. The top GPA quartile was dropped from the STAR demonstration sample because few in this group fail to graduate (7.2 percent of incoming students in 1999 in the top high school grade quartile

---

[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 et al. (2006) estimate that the vast majority of Georgia HOPE scholarships would have maintained the first-year target GPA of 2.0 even in absence of the program.

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% would qualify based on historical data. The $1,000 target was set so that about 20-25% were expected to qualify in the absence of a treatment effect. For a subset of SFP 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] STAR GPA targets are summarized in a chart in the appendix.[8]

Students receive 1 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 5 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 allow some students with fewer than 5 credits to be eligible for a merit scholarship while minimizing the incentive to take fewer courses, the GPA for 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 enrol for a second year at any college to be eligible for a fellowship. Fellowship cheques were sent to students in August after students registered for their second year. It turned out that all students with grades above their targets continued studying into their second year, without interruption and without changing university.

---

[7]Treated students were not told how their GPA target was chosen. If any students inquired, program operators were asked to tell them that the targets were individually set for research purposes. This occurred only once.
[8]Course grade distributions are not fixed. Average grades typically vary as much as 5 percentage points from year to year. Even large program effects would generate overall changes that are within this range. Effects on the order of half a standard deviation, for example (an increase of 6 percentage points), would raise the overall average by 1.5 percentage points (0.06*0.25). In fact, 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.

Shortly after they acknowledged receipt of program details, students in the SSP and SFSP were assigned advisors. The advisors emailed 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 in November detailing a student's grade targets for those who participated in the SFP and SFSP. 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 1656 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 basic student background variables – 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 1571 who registered for at least two courses completed this survey. The likelihood of completing the survey appears to be unrelated to treatment status.

The university in which this study was carried out is primarily a commuter school. Roughly eighty percent of students in our sample were living at home with their parent(s). Slightly less than a quarter identified this campus as their first choice for college. The majority

---

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

plan 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 4 years. Among those who entered 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 first year.

Merit scholarship programs like STAR may affect course enrolment 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

---

[10]Few students are French-speaking. Most non-native-English speakers in our sample are from South or East Asia.
[11]Columns labelled "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 labelled "All controls," include survey dummy variables for 9 education categories for mother and father's highest level of education (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 4 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.

positive effect on course difficulty is a response to the treatment (since financial incentives seem more likely to 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 the sample who was registered for at least two courses have a Fall grade. The likelihood of being in this group is unrelated to treatment assignment.


C. Consent Rates and Service Use

Students randomly assigned to STAR treatment groups were asked to sign up. Those who did not sign up were ineligible for services and fellowships. Sign-up imposed no burden or obligation on program participants beyond the receipt of reminder emails and mailings, including a biweekly email from peer advisors in the service programs. Students assigned to the control group were not sent any information about the demonstration. 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). 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, Panel B of Table 3 shows that 46 percent of men offered

---

[12]Students interviewed during our focus groups (discussed more 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 remarking or additional grades. 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% or more of the previous year's class received a grade no better than a D).

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 signing up.

The pattern of service use shows differences by treatment and sex similar to those observed in sign-up rates.  This pattern is presented 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 emailed 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 SFSP service use for women was about 53 percent.  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.

Specific service-use rates 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 emailed 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

---

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

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. In addition, there were 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.

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, not common gender. 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, among the sample of 215 STAR participants who signed up for an SSP or SFSP program.

The first column of Table 4 shows that females met with advisors 16 percentage points more often than males. The results in column 3, however, suggest that females were substantially less likely to meet with male advisors, while males were slightly less likely to meet with female advisors. Students were also more likely to email a same-sex advisor. While these results point to a role for gender matching in the *use* of academic services, gender differences in signup cannot be explained by the prevalence of female advisors since advisors were assigned until after students signed up. Moreover, gender differences in the use of services are greater than can be explained solely by gender-advisor interactions after assignment.

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 non-compliance.  For example, some of those offered the fellowship program were ineligible for fellowships because they did not sign up.  Likewise, those who did not sign up for the SSP and SFSP could not use services and were not contacted by Peer Advisors.  This tends to reduce 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 Bloom (1984):

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

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 an instrument for $P_i$.  The result is a covariate-weighted average effect of treatment on the treated (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 emails 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'

emails 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 (see 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 represents 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]

---

[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

Because both of the fellowship-carrying treatments produce substantial and significant effects on Fall grades, we report estimates from a models that pools 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.=.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 alone. The average fellowship effect (i.e., combining the SFP and SFSP groups using a single any-SFP dummy) is 3.1 (s.e.=.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 first-year, the SFP effects on women had faded. The estimated SFP effect on first-year GPA is .086 (s.e.=.084), which is roughly the same as the SSP effect, also insignificant at .12 (s.e.=.082). On the other hand, the effect of the SFSP treatment remains large and significant, at about .27

types, or years and therefore seems likely to be due to chance. It's also worth noting that the treatment effects reported here differ somewhat from those in our initial December 2006 NBER working paper, which 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).

(s.e.=.12) for women and .21 (s.e.=.092) overall.  Again, the overall result is driven by the effect on women.  In standard deviations units, the SFSP effect on GPAs for women is only a little smaller than the corresponding Fall grade effect (.3σ versus .35σ).  Combining both the SFP and the SFSP estimates in a single any-SFP effect generates a significant result for women of about .15 (s.e.=.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 1 and year 2.[15] Similar to the corresponding estimate in Table 5, the SFSP effect on women's first-year GPA is .24 (s.e.=.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 second year.  This estimate, at .28 (s.e.=.11), differs only slightly from the first-year effects.  Thus, the SFSP treatment appears to have generated a robust and lasting 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 permanent 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

---

[15]There are no significant differences across program groups in the fraction of students having a GPA in year 1 and the fraction of women having a GPA in year 2.  The fraction of SFP and SFSP men with GPAs in year 2 is 6.7 (s.e.=3.5) and 6.3 (3.9) percentage points higher than men in the control group.

academic probation and are at risk of suspension.[16] Many students spend time on probation. For example, 22 percent of first-years in 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 placed 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.=3.5) percentage points in the combined sample of men and women. For women, the SFSP effect on first-year probation rates is -.10 (s.e.=.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 probation rates are also persistent into second year. The second-year probation effect on SFSP women, reported in column 6 of Table 6, declines to -.097 (s.e.=.047), with an insignificant -.053 (se.=.038) effect overall.[17]

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 administrative 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 in good standing at the end of first and second years. The SSP treatment group is

---

[16]A credit is awarded for each year-long course successfully completed. A half-credit is awarded for each one-semester course successfully completed.
[17]This second-year variable codes students who have withdrawn in second year as on probation.

also more likely than the control group to be registered 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 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 is .27 (s.e.=.11) in first year and .28 (s.e.=.17) in 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.[18]

B. Effects on the Grade Distribution

The results in Table 6 suggest the STAR intervention had an impact on students who were at risk of probation or failure to maintain good academic standing. This 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, laid over the GPA distribution

---

[18]In view of the similarity of the GPA results across years, we also tried constructing more precise estimates of a common treatment effect by stacking year 1 and year 2 observations (with standard errors clustered by student). As with the single-year results, stacked estimation generated a significant treatment effect only for women in the SFSP treatment group with 10-20 percent reduction in standard errors. See Angrist, Lang and Oreopoulos (2007) for details. We also looked at a credit-weighted GPA outcome variable, constructed as GPA * (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.

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.[19]

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, which appears to shift 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 students 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, which are also indicated in the figure. This

suggests the program might have been more effective with lower fellowship targets. [20]

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. This table was constructed using stacked data from year 1 and year 2 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 the two-year sample is .264 with a standard error of .094, below both

single-year standard errors of around .11 but similar in magnitude.

---

[19]Specifically, the college GPA variable in this figure equals GPA+.3 for those in the lowest high school GPA quartile and GPA-.3 for those in the second highest high school GPA quartile.

[20]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 $1000 payment. The eligibility rates for students in the SSP, SFP, and SFSP treatment groups were similar. There was a modest increase in the likelihood that women in both fellowship groups met the standard for a $1000 award; .071 in the SFSP group, but this difference is not significantly different from zero at conventional levels. There was a marginally significant .067 gain in $5000 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).

The largest SFSP effect on the GPA distribution occurs at the .1 quantile – a finding that lines up with the GPA distribution shifts seen in the figures. But there is reasonably clear evidence of an shift at higher quantiles as well; in particular, effects at the .75 and .9 quantiles are close to the lower quartile impact of .271 (s.e.=.125). The smallest quantile effect is at the median, an estimate of .187 (s.e.=.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 a <u>conditional</u> distribution shift; that is, the effects of treatment with other included covariates held constant. The relatively flat quantile profile may mask overall asymmetry if treatment effects vary with included covariates.[21] We therefore report a set of SFSP effects using alternative sets of controls in Panel C of Table 7. This panel shows a similar pattern of effects across quantiles when the only included covariates are high school GPA and a year dummy. Interestingly, however, the quantile profile shows a relatively larger impact at the median and below when the high school GPA variable is dropped. For example, the estimated effects at the .75 and .9 quantiles become smaller without the high school GPA controls. In particular, the significant effect at the .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 – while having relatively little effect for those in the upper quartile.

---

[21] The distinction between conditional and marginal quantile treatment effects is emphasized in recent work by Firpo (2007) and Frolich and Melly (2007), who provide methods for the estimation of marginal treatment effects from conditional. 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.

C. Subgroup Effects

The results above paint a reasonably clear picture of larger SFSP achievement effects for females than for males. The results for other subgroups, however, are not as clear cut. We estimated heterogeneous effects by high school performance; expected hours of work, self-reported funding concerns; and parents' educational attainment. The additional subgroup analysis reveals some 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 SFSP females whose parents did not attend college. Specifically, the GPA and total credits earned results are significantly higher for first-college-generation SFSP females 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 with socioeconomic status. The estimated effects of SSP and SFP are fairly 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 estimated 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 ($ssp_i$, $sfp_i$, and $sfsp_i$) were used as instruments.

In addition to estimates of equation (2), we constructed 2SLS estimates of a combined any-SFP effect by estimating the parameter $\delta$ 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 $\beta$ and $\gamma$ 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 which 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, proper rescaling to take account of low SSP take-up make this investigation of independent interest for 2SLS.[22]

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. One of the most important results 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, .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 .15$\sigma$, would not.

The over-identification test statistics in Table 8 reject the hypothesis of SFP and SFSP treatment-effect equality (or come close). The difference in 2SLS estimates of $\beta$ and $\gamma$ are

---

[22]Other details related to the 2SLS procedure: The individual error component in equations (2) and (3), $\xi_i$, 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.

therefore at least marginally significantly different. Substantively, this means that the differences in intention-to-treat estimates reported in earlier tables are not simply due to differences in compliance rates.

IV. Student Reports

In order to better understand students' perception of the program and their reaction to it, we conducted open-ended interviews with students in each of the treatment groups, sampling from those who signed up. Interviewees were chosen randomly and offered $20 University Bookstore gift certificates for attending. We contacted 54 students by email 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 (2 from SFSP, 5 from SFSP, and 3 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 one male after inviting 11 other randomly selected SFP and SFSP students and offering $10 gift certificates.

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. $5,000! I definitely wasn't going to miss that." (This student received the full fellowship.) A male student we spoke with, however, said the money was not that important a motivator because "I already have enough for 1$^{st}$ and 2$^{nd}$ 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."[23]  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 the motivation to structure my time, 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 fellowship interviewees suggested that the reminders (or lack of reminders) through emails and advisor contacts mattered quite a bit.  No student said there were too many fellowship reminders or too many emails 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 simply glad to know that they

---

[23]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."

could communicate with their advisor if needed. One male SSP students noted, "University is something different and I also wanted extra help. 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 their advisor, but nevertheless felt the advisor's regular emails 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 that 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. The power of the fellowship might have been boosted by more frequent contacts and by setting lower grade targets, though this would have increased costs. On the service side, virtually all participants welcomed interaction with upper-year student advisors, even when only through email.

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 non-participation reveals a level of precision sufficient to detect theoretical service effects equal to about $.25\sigma$ 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 Year 1 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 second year. The average GPA in second year for SFSP females was about .28 higher than the GPA in the control group, a substantial and statistically significant effect. SFSP females also earned a further quarter credit more than controls in second year, so that after two years, the SFSP-control difference amounts to a little

over half a credit.  This is important because it shows that the 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.

Although limited in scope, 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.[24] 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 therefore $739, low by the standards of other social programs.[25] In future work, we plan to look at lower-cost service strategies that focus more on email-based peer advising.  Even at the current cost level, however, it's very likely that – for women - these amounts will be more than offset by future earnings gains should our estimated program effects on credits and academic standing translate into increased completion rates and consistently higher 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 across 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 less likely to report poor school enjoyment (Kleinfeld, 1998).

---

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

Goldin, Katz, and Kuziemko (2006) and Frenette and Zeman (2007) also note substantially higher college attendance and completion rates in the United States and in Canada, respectively. Jacob (2002) attributes many sex differences in college attendance and completion to differences in non-cognitive 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's 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 US 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, Anderson's (2006) evaluation of three pre-school programs suggests these program benefit girls but not boys, and the MTO evaluation (e.g., Clampet-Lundquist et al. 2006) points to benefits of subsidized housing in non-poverty 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.

## Table 1. Descriptive Statistics

| | Control Mean (1) | SSP v. Control (2) | SFP v. Control (3) | SFSP v. Control (4) | F-stat (all=control) (5) | Obs (6) |
|---|---|---|---|---|---|---|
| | | Contrasts by treatment status | | | | |
| *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) | 1656 |
| No show | 0.054 | 0.002 [0.016] | -0.030 [0.016]* | 0.020 [0.019] | 1.852 (0.136) | 1656 |
| Completed survey | 0.898 | -0.018 [0.022] | -0.010 [0.022] | -0.051 [0.028]* | 1.228 (0.298) | 1656 |
| *Student background variables* | | | | | | |
| Female | 0.574 | -0.006 [0.036] | 0.029 [0.035] | -0.005 [0.045] | 0.272 (0.845) | 1571 |
| High school GPA | 78.657 {4.220} | 0.170 [0.308] | 0.238 [0.304] | -0.018 [0.384] | 0.276 (0.843) | 1571 |
| Age | 18.291 {0.616} | -0.054 [0.045] | -0.033 [0.044] | 0.026 [0.056] | 0.752 (0.521) | 1571 |
| Mother tongue is English | 0.700 | 0.017 [0.033] | 0.009 [0.033] | 0.049 [0.041] | 0.495 (0.686) | 1571 |
| *Survey response variables* | | | | | | |
| Lives at home | 0.811 | -0.040 [0.030] | 0.009 [0.030] | -0.004 [0.038] | 0.685 (0.561) | 1431 |
| At first choice school | 0.243 | 0.024 [0.034] | 0.060 [0.033]* | 0.047 [0.042] | 1.362 (0.253) | 1430 |
| Plans to work while in school | 0.777 | 0.031 [0.032] | -0.066 [0.031]** | 0.037 [0.040] | 2.541 (0.055) | 1431 |
| Mother a high school graduate | 0.868 | 0.015 [0.026] | -0.021 [0.026] | -0.045 [0.033] | 1.040 (0.374) | 1431 |
| Mother a college gradate | 0.358 | 0.053 [0.037] | -0.020 [0.036] | -0.052 [0.046] | 1.487 (0.216) | 1431 |
| Father a high school graduate | 0.839 | 0.025 [0.028] | 0.008 [0.027] | -0.017 [0.035] | 0.416 (0.741) | 1431 |
| Father a college graduate | 0.451 | 0.021 [0.038] | -0.001 [0.037] | -0.024 [0.048] | 0.216 (0.885) | 1431 |
| Rarely puts of studying for tests | 0.208 | 0.031 [0.032] | 0.031 [0.031] | 0.107 [0.040]*** | 2.534 (0.055) | 1431 |
| Never puts off studying for tests | 0.056 | -0.019 [0.016] | -0.016 [0.016] | -0.032 [0.021] | 1.206 (0.306) | 1431 |
| Wants more than a BA | 0.556 | 0.052 [0.038] | -0.029 [0.037] | 0.073 [0.048] | (1.752) (0.155) | 1431 |
| Intends to finish in 4 years | 0.821 | -0.008 [0.030] | -0.006 [0.029] | -0.063 [0.037]* | (0.942) (0.419) | 1431 |

Notes: Standard deviations are shown in braces in column 1. Standard errors are reported in brackets in columns 2-4. P-values for the F-tests in the last column are reported in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2. Selection Effects

| | Number of credits attempted | | Number of math and science credits attempted | | Has fall grades | |
|---|---|---|---|---|---|---|
| | Basic | All controls | Basic | All controls | Basic | All controls |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. All** | | | | | | |
| *Control group mean* | 4.049 | | 1.095 | | 0.893 | |
| | (0.893) | | (1.206) | | (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 | 1571 | 1431 | 1571 | 1431 | 1571 | 1431 |
| **B. Men** | | | | | | |
| Control group mean | 3.964 | | 1.159 | | 0.914 | 0.000 |
| | (0.944) | | (1.240) | | (0.281) | (0.000) |
| 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 |
| **C. Women** | | | | | | |
| *Control group mean* | 4.112 | | 1.047 | | 0.877 | 0.000 |
| | (0.848) | | (1.179) | | (0.328) | (0.000) |
| 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. 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 10%; ** significant at 5%; *** significant at 1%

## 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 | All controls | Basic controls | All controls | Basic controls | All controls | Basic controls | All controls |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 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 | 1571 | 1431 | 1571 | 1431 | 1571 | 1431 | 1571 | 1431 |
| 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 |
| 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. 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 10%; ** significant at 5%; *** significant at 1%

## Table 4: Analysis of Gender-Adviser Interaction

| | Met with Advisor | Contacted Advisor (including email) | | Met with Advisor | Contacted Advisor (including email) |
|---|---|---|---|---|---|
| | (1) | (2) | | (3) | (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.

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

| | SFP by type | | | Any SFP | | |
|---|---|---|---|---|---|---|
| | All | Men | Women | All | Men | Women |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | | | 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 | 1255 | 526 | 729 | 1255 | 526 | 729 |
| | | | 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 | 1255 | 526 | 729 | 1255 | 526 | 729 |

Notes: The table reports regression estimates of treatment effects on the dependent variables indicated 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 10%; ** significant at 5%; *** significant at 1%

Table 6. Treatment Effects on First and Second Year Outcomes

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | All | Men | Women | All | Men | Women |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 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 | 1399 | 577 | 822 | 1241 | 521 | 720 |
| 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 | 1418 | 590 | 828 | 1418 | 590 | 828 |

Continued.

Table 6. Treatment Effects on First and Second Year Outcomes, Continued

| | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|
| | All | Men | Women | All | Men | Women |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 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 | 1418 | 590 | 828 | 1418 | 590 | 828 |
| 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 | 1418 | 590 | 828 | 1418 | 590 | 828 |

Notes: The table reports regression estimates of treatment effects on the dependent variables indicated 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 first year and probation or withdrawal in second year. The credits earned and good standing variables are zero in second year for those who withdrew. SFSP treatment effects for males and females are significantly different at the 5 percent level or better for year 2 GPA, credits earned in year 1, and weighted GPA in both years. The difference for year 2 credits earned is significant at the 10 percent level.

* significant at 10%; ** significant at 5%; *** significant at 1%

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

| Quantile | OLS | Quantile Regressions | | | | |
|---|---|---|---|---|---|---|
| | | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| A. Males | | | | | | |
| *Control mean/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 | | | 1098 | | | |
| B. Females | | | | | | |
| *Control mean/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 | | | 1542 | | | |
| 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 | | | 1542 | | | |

Notes: Standard errors are reported in brackets, standard deviations are in parentheses. 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 10%; ** significant at 5%; *** significant at 1%

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 | | 0.254 | | 2.404 | |
| | (0.907) | | (0.436) | | (1.239) | |
| SSP sign-up | 0.071 | 0.068 | 0.004 | 0.005 | 0.049 | 0.046 |
| | [0.126] | [0.124] | [0.053] | [0.053] | [0.173] | [0.171] |
| SFP sign-up | 0.041 | | -0.033 | | 0.034 | |
| | [0.081] | | [0.033] | | [0.104] | |
| SFSP sign-up | 0.315 | | -0.117 | | 0.327 | |
| | [0.112]*** | | [0.043]*** | | [0.148]** | |
| SFP (any) sign-up | | 0.125 | | -0.058 | | 0.121 |
| | | [0.070]* | | [0.029]** | | [0.091] |
| Overid test $\chi^2(1)$ | | 4.487 | | 2.900 | | 3.211 |
| | | (0.034) | | (0.089) | | (0.073) |
| Observations | 1542 | 1542 | 1656 | 1656 | 1656 | 1656 |

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 for 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.
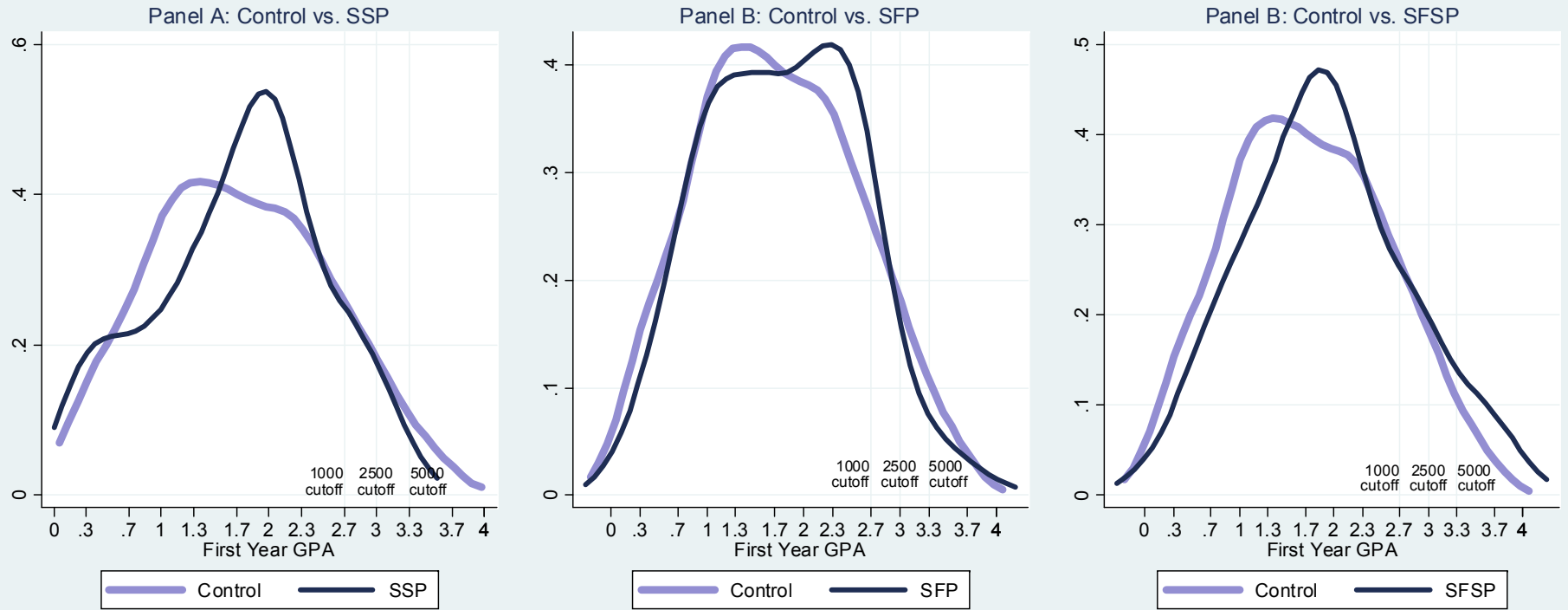
* significant at 10%; ** significant at 5%; *** significant at 1%

# Figure 1a. Males' Normalized First-year GPA
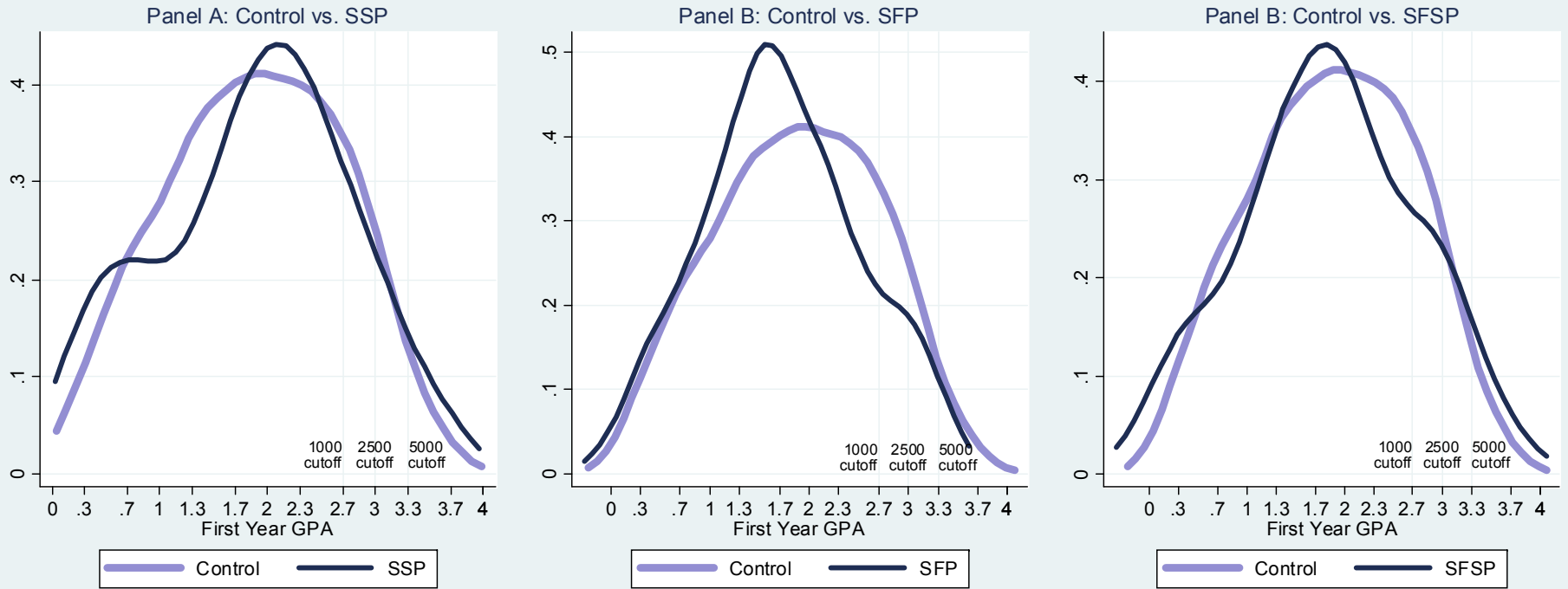


Notes: These figures plot the smoothed kernel densities of first year GPA.
The K-S p-value is a test for equality of distributions.

Figure 1b. Females' Normalized First-year GPA

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

# References

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

Angrist, Joshua (1991), "Grouped-Data Estimation and Testing in Simple Labor-Supply Models," *Journal of Econometrics* 47, 243-266.

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 D., Daniel Lang, and, Philip Oreopoulos (2006), " Lead Them to Water and Pay Them to Drink: An Experiment with Services and Incentives for College Achievement," NBER Working Paper no. 12790, December.

Angrist, Joshua D., Daniel Lang, and, Philip Oreopoulos (2007), "Incentives and Services for College Achievement: Evidence from a Randomized Trial," IZA Discussion Paper No. 3134, October.

Angrist, Joshua and Victor Lavy (2002), "The Effect of High School Matriculation Awards: Evidence from Randomized Trials," NBER Working Paper no. 9389, December.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King and Michael Kremer (2002), "Vouchers for Private Schooling in Colombia: Evidence from Randomized Natural Experiments," *American Economic Review*, 92(5), pp. 1535-1558.

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," Mimeo, University of Missouri-Kansas City.

Ashworth, K. et al. (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), January.

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), pp. 9-18.

Bettinger, Eric and Bridget Terry Long (2005), "Addressing the Needs of Under-Prepared Students in Higher Education: Does College Remediation Work?" NBER Working Paper no. 11325, May.

Bloom, Dan and Colleen Sommo (2005), "Building Learning Communities: Early Results from the Opening Doors Demonstration at Kingsborough Community College," MDRC Report, June.

Bloom, H. S. (1984) "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, 8 (2).

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.

Card, David (1995), "Earnings, Schooling, and Ability Revisited," in Solomon Polachek, ed., *Research in Labor Economics*, 14.

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

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 IRS Working Paper no. 509, March.

Congos, D. 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), pp. 159-170.

Consortium for Student Retention Data Exchange (2004), *2003-04 CSRDE report: The retention and graduation 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), pp. 761-786.

DesJardins, S. and Brian McCall (2006), "The Impact of the Gates Millennium Scholarship Program on the Retention and Academic Performance of Minority Students: A Regression-Discontinuity Analysis," Carlson School of Management, University of Minnesota, Mimeo, October.

Dominitz, Jeff and Charles F. Manski (2000), "Using Expectations Data to Study Subjective Income Expectations," *Journal of the American Statistical Association*, 92(439), pp. 855-867.

Dynarski, Susan (2002), "The Consequences of Merit Aid," NBER Working Paper no. 9400, December.

Dynarski, Susan (2005), "Building the Stock of College-Educated Labor," NBER Working Paper no. 11604, September.

Firpo, Sergio (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, Vol. 75, No. 1, pp. 259-276.

Frederick, S., Lowenstein, G., and T. O'Donoghue (2002), "Time Discounting and Time Preferences: A Critical Review", *Journal of Economic Literature*, XL, 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 No. 303, September.

Frölich, Markus, and Blaise Melly (2008), "Unconditional Quantile Treatment Effects Under Endogeneity," IZA Discussion Paper No. 3288, January.

Fudenberg, Drew, and David K. Levine (2006), "A Dual Self Model of Impulse Control," American Economic Review, Vol. 96, No. 5 (December), pp. 1339-1476.

Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino and Enrico Rettore (2007), "College cost and time to obtain a degree: Evidence from tuition discontinuities," NBER Working Paper no. 12863, January.

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*, Vol. 20, No. 4, pp. 133-156.

Goodlad, J.I. 2004. A place called school: Prospects for the future. 20th anniversary ed. New York: McGraw-Hill.

Grossman, J. B., and J. P. Tierney (1998), "Does Mentoring Work? An Impact Study of the Big Brothers Big Sisters program," *Evaluation Review,* 22(3), pp. 402–425.

Habley, W.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," *The Journal of College Student Development*, 44(2), pp. 250-259.

Imbens, Guido W. and Joshua D. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica,* 62(2), 467-475.

Jackson, C. Kirabo (2007), "A Little Now for a Lot Later: A Look at the Texas Advanced Placement Program," mimeo, Harvard University.

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*, Vol. 21, pp. 589-598.

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

Kleinfeld, J. (1998), "The Myth That Schools Shortchange Girls: Social Science in the Service of Deception," Women's Freedom Network, Washington, D.C. ERIC (Education Research Information Clearinghouse) document number ED 423 210.

Kremer Michael, Edward Miguel and Rebecca Thornton (2004), "Incentives to Learn," CID Working Paper no. 109, October.

Laibson, David. "Golden Eggs and Hyperbolic Discounting," Quarterly Journal of Economics, Vol. 112(2), 1997, pp. 443-77.

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

Leuven, E., H. Oosterbeek and B. van der Klaauw (2003), "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment," CEPR Discussion Paper no. 3921.

Lotkowski, V. A., S. B. Robbins and R. J. Noeth, (2004), "The Role of Academic and Non-academic Factors in Improving College Retention," ACT Policy Report (ERIC Document Reproduction Service no. ED485476).

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

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

National Center for Education Statistics (2003), *Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000,* Washington DC: Department of Education.

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

Ogden, P., D. Thompson, H. A. Russell and C. 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, forthcoming.

Pascarella, E. T. and P. 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. "The adolescent brain and age-related behavioral manifestations," Neuroscience and Biobehavioral Reviews, 24 (2000), pp. 417-463.

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

Wyckoff, S. (1998), "Retention Theories in Higher Education: Implications for Institutional Practice," *Recruitment and Retention in Higher Education*, 12(2), pp. 2-7.

**Appendix: Student Fellowship Program Award Schedule**

| Previous High School Grade Avg. Quartile | Award Thresholds | | |
|---|---|---|---|
| | 1000 | 2500 | 5000 |
| 0 – 25$^{th}$ percentile | 2.3 (C+) | 2.7 (B-) | 3.0 (B) |
| 25 – 50$^{th}$ percentile | 2.7 (B-) | 3.0 (B) | 3.3 (B+) |
| 50$^{th}$ – 75$^{th}$ percentile | 3.0 (B) | 3.3 (B+) | 3.7 (A-) |

Notes: Eligibility was determined by the student's best 4 courses. Half of SFP/SFSP participants were offered the 2500 award.