NBER WORKING PAPER SERIES

LEAD THEM TO WATER AND PAY THEM TO DRINK: AN EXPERIMENT WITH SERVICES AND INCENTIVES FOR COLLEGE ACHIEVEMENT

Joshua Angrist
Daniel Lang
Philip Oreopoulos

Working Paper 12790
http://www.nber.org/papers/w12790

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2006

This project was sponsored by the Canada Millennium Scholarship Foundation. Special thanks go to Yan Tam-Seguin and Daniela Rukavina, as well as to other staff and administrators who assisted with the project, and to Cynthia Kinnan for exceptional research assistance. We benefited from helpful comments from seminar participants at the National Bureau of Economic Research Summer Institute, Stanford University, University of California, Santa Cruz, the University of Chicago, the University of Toronto, Williams College, Harvard University, the University of Houston, Rice University, Texas A&M, Statistics Canada, the London School of Economics, MIT, and the Stanford Institute Theoretical Conference (SITE). The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research or the Canada Millenium Scholarship Foundation.

Lead Them to Water and Pay Them to Drink: An Experiment with Services and Incentives for College Achievement

Joshua Angrist, Daniel Lang, and Philip Oreopoulos

NBER Working Paper No. 12790
December 2006
JEL No. I22,I28,J24

ABSTRACT

High rates of attrition, delayed completion, and poor achievement are growing concerns at colleges and universities in North America. This paper reports on a randomized field experiment involving two strategies designed to improve these outcomes among first-year undergraduates at a large Canadian university. One treatment group was offered peer advising and organized study group services. Another was offered substantial merit-scholarships for solid, but not necessarily top, first year grades. A third treatment group combined both interventions. Service take-up rates were much higher for students offered both services and scholarships than for those offered services alone. Females also used services more than males. No program had an effect on grades for males. However, first-term grades were significantly higher for females in the two scholarship treatment groups. These effects faded somewhat by year's end, but remain significant for females who planned to take enough courses to qualify for a scholarship. There also appears to have been an effect on retention for females offered both scholarships and services. This effect is large enough to generate an overall increase in retention. On balance, the results suggest that a combination of services and incentives is more promising than either alone.

Joshua Angrist
Department of Economics
MIT, E52-353
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER
angrist@mit.edu

Philip Oreopoulos
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON M5S 3G7
Canada
and NBER
oreo@economics.utoronto.ca

Daniel Lang
The Ontario Institute for Studies in Education
University of Toronto
252 Bloor Street West
Toronto, ON M5S 1V6
dlang@oise.utoronto.ca

I. Introduction

Recent years have seen growing interest in policy strategies designed to increase college attendance and completion, especially for low-income students. Major efforts to increase enrolment include merit- and need-based aid, tax deferral programs, tuition subsidies, part-time employment assistance, and improvements to infrastructure. These expenses are justified in part by empirical evidence which suggests that there is a substantial economic return to a college education (see, e.g., Kane and Rouse, 1995).

In addition to the obvious necessity of starting college, an important part of the post-secondary education production function is student achievement and retention. Many students perform poorly and take much longer to attain a degree than the nominal completion time. First-year students are especially likely to struggle. Nearly one-third of first-year college students in the U.S. take 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 and about two in five leave within six years without a degree. Moreover, fewer than half of Black and Hispanic students and students attending colleges with a predominantly part-time student body graduate within six years (Consortium for Student Retention Data Exchange, 2004).[1]

Social scientists often view schooling decisions as outcomes of an optimization problem. In this context, the decision to leave college or not study seriously may be an individually rational response to new information about costs and benefits. Nevertheless, policy-makers

[1] Pantages and Creedon (1978) summarize research on college retention from 1950 to 1975, and Peltier, Laden, and Matranga (1999) and Lotkowski, Robbins, and Noeth (2004) review more recent research. Interestingly, the three articles report consistently high college attrition rate with little downward trend over time. The average six-year graduation rate, for example, among students that entered a 2 or 4 year college program, was about 40 percent in 1957 (Pantages and Creedon (1978), 40 percent between 1985 and 1996 (Peltier, Laden, and Matranga, and 40 percent in 2000 (Lotkowski, Robbins, and Noeth (2004).

typically see dropout behaviour and low achievement as undesirable and college administrators invest considerable time and money in an effort to increase retention, speed degree completion, and raise achievement. Possible rationales for this include students' failure to account for sheepskin effects on wages (Jaeger and Page, 1995) or an economic return to within-college investment in learning (Loury and Garman, 1995). Moreover, at heavily-subsidized public institutions, post-secondary education costs tax-payers more when it takes longer.

Motivated by the view that achievement problems reflect a weak academic background, the traditional response to retention and achievement problems has been an array of academic service strategies (Barefoot, 2004). For example, most North American institutions offer subject-specific tutorials and one-on-one tutorial support, extra drop-in hours, and remedial courses. Sometimes these services are combined with psychological support services and efforts to boost general time-management skills, motivate students, and facilitate integration into the college social environment (Tinto, 1993; Goodlad, 2004). Although there is some observational and anecdotal evidence suggesting students who make use of these services reap benefits, the observed relation between outcomes and support services need not be causal.

Merit-based aid also has a long history in the post-secondary context, but traditional programs, like US National Merit awards and Canadian Excellence Awards, have focused on a relatively small number of very high achievers. A recent development in the scholarship field is an attempt to use financial awards and incentives to motivate good but not spectacular students.[2] Examples of this effort include state tuition waivers for students who maintain a B-average and other problems modeled on Georgia's HOPE program. As Dynarski (2005) notes, these are not elite programs. For example, nearly 60% of Georgia high school graduates qualify for a scholarship (assuming they go to college). A number of quasi-experimental evaluations suggest

[2]The National merit program awards 8200 scholarships to students selected from 1.4 million PSAT-takers.

these new scholarship programs boost college attendance and completion (Dynarski, 2003, 2005), but the evidence is mixed and the costs of these programs are large (Cornwell et al. 2006).[3]

To the best of our knowledge neither academic support strategies nor financial incentives have been the subject of large-scale evaluations using a random-assignment research design in a college setting. The purpose of this paper is to report on a large randomized field experiment designed to assess major strategies now being used to improve college retention and academic achievement. Approximately 1,600 first-year students participated in the Student Achievement and Retention Project (STAR) at a large Canadian university. In American terms, this institution can be thought of as a large state school, with tuition heavily subsidized.

The STAR demonstration project involved most of the entering class at one of the university's satellite Arts and Science campuses. The satellite campus is of special interest in this context since achievement and retention are more of a problem than on the main campus. 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 that provides critical thinking strategies for performing well in a particular course. A second group was offered substantial cash awards – up to the equivalent of a full year's tuition – for meeting a target GPA. Finally, a third treatment group was offered services and incentives, a combination that has not been looked at previously using any sort of research design.

[3]The largest US aid program besides veterans benefits is the Pell grant program. Bettinger (2004) finds that Pell grants reduce dropout rates.

The rest of the paper is organized as follows. The next section reviews some theoretical background and previous literature on related post-secondary interventions. Section III describes the STAR demonstration. Section IV discusses the results and Section V reports on our discussion with focus groups in an attempt to understand the findings.

Overall, the effects of the STAR intervention were small. A closer look, however, reveals a number of important effects on some females. Females in both fellowship groups had markedly better Fall term grades and somewhat better grades at year's end, though the initial boost faded considerably. The effects on females come from the subgroup planning to take at least four courses, the minimum load required to qualify for a fellowship. Another important result is that the year-end grade and retention effects were stronger in the group that combined both services and incentives. The combination of incentives and services also generated much higher service use than the offer of services alone. The results therefore suggest that a combination of services and incentives is more promising than either intervention alone.

II. Background and Context

The benchmark economic model of schooling-as-human-capital treats educational attainment as the outcome of an optimization problem solved by equating marginal costs and benefits. This framework allows for heterogeneous costs and benefits, thereby generating a distribution of schooling choices even among observationally similar individuals (see, for example, Card, 1995). In this framework, new information may make some students update their assessment of costs or benefits and therefore leave school. For example, students may discover that the college workload is higher than they anticipated or that they dislike studying college-level material. Moreover, some studies suggest that the economic returns to a partially

completed degree are not substantially below the returns to degree completion in per-year terms (Kane and Rouse, 1995). Viewed in this way, it is not clear why the decision to leave school or sub-par academic performance should be of concern to economists or policy-makers.

In practice, a number of considerations suggest we should not be sanguine about poor performance and college dropouts. First, additional support programs offered to students may increase educational benefits or lower costs. Moreover, students may incorrectly gauge the economic consequences of poor performance or leaving school (Dominitz and Manski, 2000). Some observers also see young people as having very high discount rates that might later change, so that the choices they make are not time-consistent (Oreopoulos, 2006). In this case, students' lifetime welfare might be enhanced by efforts that increase the motivation to do well in school.

The traditional approach to retention and completion focuses on fostering academic skill. College students clearly run into trouble when they are poorly prepared for college work. Proxy variables for academic background, such as high school GPA or standardized entrance test scores, are the best single predictors of first-year college performance and attrition (Lotkowski, Robbins, and Noeth, 2004). Aware of this fact, many institutions offer an array of services, including remedial courses, academic advising, orientation classes, content-based tutoring, and writing workshops. A service strategy known as Supplemental Instruction, which plays a role in our evaluation, tries to promote critical thinking and reasoning skills.

Non-experimental evidence on the effectiveness of student services is mixed (see surveys, for example, by Bailey and Alfonso, 2005, Pascarella and Terenzini, 1991, Lotkowski, Robbins, and Noeth, 2004, and Wyckoff, 1998). More rigorous studies with experimental and quasi-experimental designs, mostly for high school students, paint a more promising picture. At the high school level, Tierney and Grossman (1998) examine a program that randomly assigned

Big Brother/Big Sister applicants to either a matched advisor or a waiting list where they remained on for at least 18 months. Youth matched to advisors were substantially less likely to use drugs and skip school. Lavy and Schlosser (2006) find positive effects of a remediation program to help weak students pass a high school matriculation exam. At the college level, Bettinger and Long (2005) report gains in retention as a result of remedial freshmen courses. Bloom and Sommo (2005) analyze early outcomes from a program that sorted freshman college students into small groups taking the same first year classes. Students randomly assigned into these 'learning communities' were more likely to pass required English courses than a control group, but second-year retention rates were unaffected. As far as we know, there have been no other randomized evaluations of college support services.

Merit scholarships have grown substantially in recent years, in both absolute and relative terms. Recent programs introduced by several U.S. states differ from previous, more private-based merit aid in that they offer more broad based rewards to students with solid, though not necessarily exemplary academic records. The Arkansas and Georgia merit scholarships for students at public universities pay students tuition as long as they maintain a GPA of B or better. These programs are partly an effort to attract better students to public institutions. But they are also motivated by the view that merit-aid increases interest in school and makes students more willing to develop good study habits.

A few recent studies look at the impact of financial incentives on the performance of college students. Garibaldi, *et al* (2006) find that Italian university students finish school more quickly when tuition is *increased* for those who run past the nominal completion time, while Dynarski (2005) finds that the Georgia and Arkansas merit-based aid programs increase enrolment rates by 4 percentage points and completion rates by 2 percentage points. Also

relevant is Leuven, Oosterbeek and van der Klaauw (2005), who conducted an experiment with incentives for a small sample at the University of Amsterdam. They report mixed effects. DesJardins and McCall (2006) have looked at early outcomes in an evaluation of the Gates Foundation effort to boost college achievement among minority students. Finally, Brock and Richburg-Hayes (2006) present early results from an experiment that offered $1,000 to low-income parents attending community college for maintaining at least a half-course load in first year and another $1,000 for maintaining a 2.0 (or C) grade average. Semesters enrolled and credits completed were both significantly higher for students offered this program.

Other evidence on incentives for academic performance comes from pre-college populations. Ashworth et al. (2002) explore the impact from providing stipends for high school students to stay in school in a non-experimental evaluation. 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. 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, especially for girls.

To the best of our knowledge, STAR is the first randomized evaluation of a merit-aid program with scholarship amounts and grade targets that closely resemble actual state-sponsored merit-based aid programs. Our study is also the first to examine a program that simultaneously targets academic skill and motivation. Tinto's (1993) pioneering work on retention emphasizes this interaction.

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

A. Study Design

The STAR demonstration involved 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 known as 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. They 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 grade in first year. Participants in the merit scholarship program received $5,000 cash, almost exactly the same as a year's tuition, for a grade average of B (a GPA of 2.0) or higher, and $1,000 for a C+ (a GPA of 1.7) or better.[4] To be eligible for a fellowship, students had to take at least 4 courses per term and register to attend the second year of their program (a full load, required to complete a degree program in four years, is 5 courses per year). In the 2003-4 school year, 7-8 percent of registered students met the standard for a $5000 award, while 26-28 percent met the standard for a $1000 award. As it turns out, however, award rates in our cohort were somewhat lower.

[4] 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, roughly 0.90 US.

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 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.[5]

The SSP strategy was motivated in part by the view that retention is strongly influenced by a student's interaction with individuals who take an interest in their welfare (Habley, 2004). Several universities assign first year students to upper-class peers or faculty advisors who provide academic support. Wyckoff (1998) suggests these informal and formal interactions increase persistence. Few colleges, however, have structured as extensive a mentoring program as was offered through the SSP. Peer advisors in the STAR program were hired based on exceptional social and academic skills. They participated in a 3-day training course as well as ongoing training feedback sessions with supervisors. The advisors were more proactive than those in typical mentor programs in that they emailed at least once every two weeks to remind advisees of their 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, encouraging advisees to use these services and to go to tutorials and faculty office hours. Advisors were also trained to identify circumstances that called for more professional help and to make appropriate referrals.

[5] The fraction treated was small relative to the total first year population. 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. The STAR demonstration was not advertised to the control group and we received few inquiries from controls or other non-program students about the program. Some treated students discussed their offer with schoolmates, but no one interviewed said that the program was a source of discussion throughout the year, or a concern for schoolmates they mentioned the program to.

The second component of the SSP was the availability 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. For example, 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 and are a commonly utilized student service in North America (e.g. Lotkowski, Robbins, Noeth, 2004). A number of studies suggests students who participate in FSG-style supplemental instruction outperform non-participating peers (Congos and Schoeps, 2003, Hensen and Shelley, 2003, Ogden, Thompson, and Russell, 2003). The STAR demonstration offered FSGs for approximately half of the largest first year courses.[6]

The 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.[7] Grade targets were therefore set as a function of high school GPA. The top GPA quartile was dropped from the entire 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 had not graduated by 2006, compared to 35.3 percent of students in the other quartiles). For each remaining quartile, the $5,000 target was set so that without the intervention, about 5 to

[6] 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.

[7] Dynarski (2005) and Mustard 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.

10% would reach it based on historical data. The $1,000 target was set so that about 20-25% were expected to qualify it in the absence of a treatment effect. For a subset of SFP students, we also offered an intermediate target of $2500. 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.[8] The exact targets appear in a chart in the appendix.[9]

Students receive 1 credit for taking a two-semester (Fall and Spring) course and half a credit 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 their 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.

Shortly after consenting to participate, students in the SSP and SFSP were assigned advisors. The advisors emailed participants in an effort to set up an initial meeting. FSG times

[8] 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.

[9] 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 average for control students fell 3 percentage points relative to students in the same high school GPA quartile from the previous two years.

and locations were announced often.[10] After the first semester, bookstore gift certificates were offered to those who attended FSGs or 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. Evaluation Framework

In practice, we cannot compel students to use services or even require the acceptance of fellowships. We therefore used an intention-to-treat design where students in one of the randomly selected treatment groups we made aware of the program available to them, while controls had no knowledge of the experiment other than what they might have heard from friends or classmates (there was also a STAR office, where anyone was free to inquire). Because the offer of treatment was randomly assigned, a simple comparison of means provides an unbiased estimate of the effect of the offer of fellowships or services. We also reported regression-estimates of intention-to-treat effects, using models that control for covariates available in administrative and survey data. The regression estimates provide a check on the unconfoundedness of the experimental random assignment, and may generate an efficiency gain.

The bulk of the estimates reported below are intention-to-treat effects that make no adjustment for non-compliance. In cases where program effects are zero, a zero intention-to-treat effect implies a zero effect on participants. More generally, however, the intention-to-treat analysis dilutes non-zero program effects. For example, about 10 percent of those offered the fellowship program did not consent to participate; these students were not eligible for fellowships even if they met SFP grade targets and therefore should be unaffected by the

[10] After the first semester, we also offered up to $50 university bookstore gift certificates for advisor and FSG contact to encourage more participation.

intervention. Likewise, those offered services through the SSP or SFSP need not have consented. We therefore report estimates that use the offer of services as an instrumental variable (IV) for program participation (in this case, consent) for a subset of samples and outcomes. This generates an estimate of the effect of treatment on program participants.

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 setting is the adjustment to intention-to-treat effects originally proposed by Bloom (1984)

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

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-as-consent and compliance-as-service-use. On the other hand, the availability of services and the interest in students shown by peer advisors (who emailed participants biweekly) is an intervention to which all SSP and SFSP participants were exposed, whether or not they actively sought services. In focus groups, treated students reported that they took note of this interest. We therefore make no adjustment for the difference between consent and usage in the 2SLS analysis.

C. Student and School Background

Almost all of the 1656 students who were selected for random assignment in August of 2005 were registered for class that Fall. This can be seen in Table 1, which reports means and differences in means by treatment group for key administrative and background variables. In July, prior to treatment selection, we surveyed all incoming first year students. Almost 90

percent of those who ended up in our sample completed the background survey.[11] About 84 percent have Fall grades, meaning they completed one or more Fall-semester courses. For these students we have intermediate grade outcomes halfway into the school year. The average planned course load at pre-registration was about 4 courses, less than a full load, but enough to qualify for STAR fellowships, according to program rules. There are no significant differences by treatment status in these pre-treatment variables, though we look again at selection issues relating to post-treatment variables in Table 2.

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 plan to work at least part-time while in college (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.[12] The students' parents, however, are reasonably well-educated; many have college degrees (though it should be noted that these are student-reported schooling measures). Most respondents claim to put a priority on high grades. Interestingly, 55 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. We tried to minimize this behaviour by contacting

[11] The high response was obtained after first making the survey online, sending a letter by the university president encouraging students to participate, offering a chance to win a laptop, several email reminders, and, finally, calling nonresponders.

[12] Few students are French-speaking. Most of the non-English speakers in our sample are from South or East Asia.

treated students for the first time only after they had completed their initial course selection. IN our setting, it would have been difficult for students to change courses later because of capacity and scheduling constraints.

Tables 2a reports treatment effects on the likelihood that students registered in the Fall and completed Fall or full-year courses for credit. Table 2b shows effects on students' course load or the number of math and science credits completed (these courses are considered more difficult). The estimates in these tables indicate that the STAR demonstration did not affect initial course load or selection, or the overall propensity to register and complete particular types of courses.

D. Consent Rates and Service Use

Students randomly assigned to STAR Demonstration treatment groups were asked to sign statements of informed consent as a condition of eligibility for services or fellowships. Informed consent imposed no burden or obligation on program participants beyond receipt of reminder emails and mailings, including unsolicited biweekly email from peer advisors in the service programs. Students assigned to the control group were tracked with administrative data and were not sent any information about the demonstration.[13] Consent serves as an indicator of student awareness and interest. About half of those randomly assigned to receive services in the SSP consented, a statistic reported in Panel A of Table 3a (columns 1 and 2). The table shows treatment-control differences and coefficient estimates from regression models that include two

[13] Members of the control group who inquired about the program were given general information. We received few such inquires.

sets of covariates.[14] Consent rates were much higher for the SFP than for the SSP, about 85 percent. SFSP consent rates were about 74 percent.

Females in each of the three treatment groups were much more likely than males to consent to participate in STAR. For example, column B of Table 3a shows that 46 percent of males offered the SSP consented, in contrast with 62 percent of females, a statistic reported in the same column in Panel C. Most students consented to the SFP, but a gap by sex remains, with 91 percent of females and 81 percent of males having consented. Similarly, when offered both services and fellowships in the SFSP, 84 percent of females and 71 percent of males consented.

The pattern of service use shows differences by treatment arm and sex similar to those observed in consent rates. In particular, service use was higher in the SFSP than the SSP (i.e., when services were offered with fellowships versus services alone), in both the Fall and Spring terms. Females were also much more likely to use services than males, even among those who consented. For example, 16 percent of all students offered services in the SSP used either peer mentoring or supplemental instruction in Fall and Spring terms, but service use was close to 30 percent in the SFSP. Fall term service use by females was 35 percent, while Fall term service use for males was about 20 percent. These estimates appear in columns 3-6 of Table 3a. The fact that service use rates were fairly stable across terms is important. Also, the fact that service use was higher in the SFSP than in the SSP in both terms suggests that the opportunity to win a fellowship motivated students to use services throughout the year. This may signal increased effort throughout the year as well.

[14] In this table and those discussed below, 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 at pre-registration. These variables come from administrative data. Columns labelled "All controls," add the responses to five survey questions, whether this project university was the subject's first-choice, hours/week of study in high school, hours/week of planned market work at university, and parents' education.

Overall service-use rates were somewhat higher than the term-specific rates reported in Table 3a. This can be seen in columns 1 and 2 of Table 3b, which reports the probability that SSP and SFSP students used services at any time during the year. The highest rate of use - almost 50 percent - was for females in the SFSP. The lowest was for males in the SSP, at 21 percent. Table 3b also shows more use of the peer mentoring service than of supplemental instruction offered through FSGs. About 10 percent of males and females in the SSP attended at least one FSG (most of those who attended once, attended more than once), while 16 percent of males and 26 percent of females met or emailed a peer advisor (excluding advisor-initiated contacts). Usage rates for both types of services were higher in the SFSP than the SSP, with 45 percent of females in the SFSP having contacted a peer advisor and 14 percent having attended an FSG.

Take-up rates for the FSG service were lower than the rates we aspired to, and probably diluted somewhat by our inability to offer FSGs in every course in which STAR participants were enrolled (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. Finally, there were unavoidable scheduling conflicts. On the other hand, students made considerable use of the advising services and reported in our follow-up focus groups that Peer Advisors were a valuable resource.

Models allowing for interactions with students' academic background and pre-registration course-load reveal important differences in consent rates and service use. These differences are explored in Table 3c, which reports estimates of main effects and interaction terms in the model:

$$y_i = X_i'\pi + \kappa_i ssp_i + \lambda_i sfp_i + \mu_i sfsp_i + \epsilon_i, \quad (1)$$

where

$$\kappa_i = \kappa_{0i} + \kappa_{1i} lowhs_i + \kappa_{2i} load4_i$$

$$\lambda_i = \lambda_{0i} + \lambda_{1i} lowhs_i + \lambda_{2i} load4_i$$

$$\mu_i = \mu_{0i} + \mu_{1i} lowhs_i + \mu_{2i} load4_i,$$

where y_i is a dependent variable, either consent or service-use, X_i is vector of covariates, and the individual specific treatment effects allow for differential effects according to whether students are in the lowest high school GPA quartile ($lowhs_i$) and whether they pre-registered for at least 4 courses ($load4_i$). Table 3c also reports sums of main effects and interaction terms (e.g., $\kappa_{0i} + \kappa_{1i}$ and $\kappa_{0i} + \kappa_{1i} + \kappa_{2i}$).

One of the largest take-up differentials documented in Table 3c is a lower consent rate and reduced likelihood of service use by males with low GPAs. For example, the estimate in column 2 shows a .22 (s.e.=.09) reduction in SSP consent for bottom-quartile males and column 5 shows a .17 (s.e.=.-07) reduction in usage rates. The corresponding estimates for females, reported in columns 3 and 6, also show negative interactions with $lowhs_i$, though these are not significantly different from zero. Another differential documented in Table 3c is the consistently higher take-up and SSP service-use rate among students who pre-registered for at least 4 courses. Consent rates for those with $load4_i$ switched on are also elevated for those assigned to the SFP and SFSP. This is important because it suggests we should expect larger reduced-form effects in the group taking more courses.

IV. Results

We begin the analysis of achievement effects by looking at students' average grades in the Fall semester and at the end of their first year of study. The grade variables are credit-

weighted averages on a 0-100 grading scale. Fall grades provide an initial measure of program impact, though some students (about 15 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 2b). As we discuss below, however, students with Fall grades are much more likely to have signed up for a full course load.

Students assigned to the SFP and SFSP earned Fall grades about 2 points higher than the control group. This is shown in columns 1-3 of Panel A in Table 4, which report treatment effects on males and females, estimated with varying sets of controls. Most of the SFP and SFSP effects shown in the table are significantly different from zero. In contrast, SSP effects are close to zero (.2-.4), and insignificant, though they are estimated with approximately the same precision as the SFP and SFSP effects.[15]

The overall impact on Fall grades is driven entirely by large and significant effects on females. This is apparent in the comparison of Panels B and C in Table 4. For example, females assigned to the SFP earned a Fall grade almost 3 points higher than the control group, while females assigned to the SFSP earned a Fall grade about 3.5 points higher than controls. The effect of both fellowship treatments on males is much smaller, and none of the estimated fellowship effects on males is significantly different from zero. Another important result is that the estimates for females suggest the combination of services and fellowships offered in the SFSP had a larger impact than the SFP (i.e., fellowships, alone.)

By the end of first-year, the SFP effects on females' Fall semester grades had faded somewhat, but remain substantial and at least marginally significant. For example, the estimated effect of the SFP on females' first-year grades in a model with basic controls is 1.7 (s.e.=1.04),

[15] Equal precision of reduced form estimates does not imply equal precision of program-use effects. We discuss this further in the context of 1 two-stage least squares analysis, below.

while the corresponding estimate of the SFSP effect on females is 3.3 (s.e.=1.5). These estimates are from column 5 of Panel C in Table 4. This is down from effects of 2.7 and 3.5 on Fall grades in the same specification and sample (reported in column 2).

Figure 1a shows that neither the offer of services through the SSP nor the offer of a fellowship through the SFP or SFSP (in combination, we call these two treatments "Any SFP") had a significant effect on the distribution of Fall and first-year grades for males in the Fall grades sample. The smallest of the Kolmogorov-Smirnov (K-S) p-values reported at the bottom of each panel in the figure is .46. On the other hand, as shown in Figure 1b, the offer of a fellowship had a marked effect on the distribution of Fall term and first-year grades for females in the Fall term grades sample. There is also at least a marginally significant distribution shift associated with the SSP, though as Table 4 shows, this did not translate into a significant mean effect (except in one specification with the full set of survey controls).

The point-wise confidence intervals for distribution treatment effects plotted in Figure 2 show where in the distribution of Fall and first-year grades the STAR interventions appears to have had an effect. The solid lines in each panel of Figure 2 were constructed from a sequence of regressions with $\{1[Y_i>c]$; c=.05, .1, .15 . . . .95$\}$ on the left hand side. There is little evidence of an effect anywhere in the grade distributions for males, except possibly a negative effect of the SFP at one point (see Panel B in Figure 2b). Figure 2c shows a pronounced effect of Any SFP on the distribution of Fall grades for females, however, with significant effects in a range around 60. The effects on first-year grades, reported in Figure 2d, are similar. Consistent with the pattern of mean effects, the distribution shift induced by the SFSP appears to be larger than the shift induced by the SFP.

The estimates in Table 4 and Figures 1 and 2 were constructed using data from the Fall grades sample; i.e., for the roughly 85 percent of students who completed one or more one-semester courses. The Fall term grades results are important because they serve to establish that the fellowship program had some effect (on females), and that the STAR research design was powerful enough to detect effects of a plausible magnitude. Moreover, the Fall grades effect on females persists through the end of first-year for the Fall grades sample.

In the full sample, program effects on first-year grades are noticeably weaker than in the Fall grades sample. This is apparent in Panel A of Table 5, which reports estimates of effects on first-year grades in the full sample. The results in columns 1-3, from a specification similar to that used to produce the estimates with basic controls in Table 4, show no significant effects on males or females. The results in columns 4-6, which report estimates from models where the SFP and SFSP effects are combined into a single Any-SFP effect (i.e., a regression on ssp_i and $anysfp_i \equiv sfp_i + sfsp_i$), are also insignificant; the effect on females in this specification is 1.4 (s.e.=.93). The difference between the full and Fall term grades samples is highlighted by Panel B, which shows results from the same specifications using the sample of students with a Fall grade. The any-SFP effect on females in the Fall term grades sample is 2.3 (s.e.=.91).

What accounts for the larger fellowship effects in the Fall grades sample? A likely explanation is that the Fall term grades sample consists almost entirely of students who pre-registered for four or more courses per term. [16] Four courses per term is the minimum required to qualify for an SFP fellowship, while five courses is considered a full load. Students taking this many courses are therefore more committed to their studies than those taking a lighter load.

[16] About half of the females pre-registered for a full load of 5 courses (430 out of 900), 412 of these have a Fall grade. 88 percent (791) of females pre-registered for 4 or more courses; 718 of these have Fall grades. Thus, most females with a Fall grade registered for at least 4 courses (718/795, just over 90 percent).

Estimates in the sub-samples of students taking or four or five courses, reported in Panels C and D of Table 5, support the notion that part of the explanation for the difference in results between the Fall term and full samples is the fact that those with a Fall grade took more courses. The estimates for females in column 3 of Panels C and D show at least marginally significant SFSP effects, though the SFP-only effect is smaller than in the Fall sample and insignificant. However, the combined Any-SFP treatment effect is marginally significant effect in the sample pre-registering for a full load of five courses.

On balance, these results suggest that while the SFP and SFSP generated an initial improvement in achievement, the full-year fellowship results were modest. There was no effect of either services or fellowships on males, while the initially strong effects for females appear to have faded. Full-year estimates for the sample with fall grades continue to show something along the lines of the Fall term results, but this may be fortuitous. On the other hand, sample-section criteria based on course load generate significant results for females offered a combination of fellowships and services.

The remainder of this section looks briefly at effects on other outcomes, focusing initially on the impact of the STAR treatments on the distribution of student GPAs. We also explore the issue of treatment-effect heterogeneity through models with interaction terms. We then turn to effects on credit units awarded at the end of first year, and retention into second year. These outcomes are two important program targets. Finally, we discuss results from a 2SLS procedure that adjusts reduced-form estimates for non-compliance.

A. Eligibility and GPA Effects

Overall, the STAR treatments failed to increase the likelihood that students earned a GPA above the targets assigned to those in the SFP and SFSP. We determined this by coding a dummy for theoretical fellowship eligibility in both the treatment and control groups (including the SSP), and using this as the dependent variable in a regression on treatment dummies and covariates. For example, a little over 16 percent of all control students finished their first-year with a GPA that qualified for a $1000 payment, but the eligibility rates for students in the SSP, SFP, and SFSP treatment groups were similar. These results are reported in the first column of Table 6.[17]

There was a modest increase in the likelihood that females in both fellowship groups met the standard for a $1000 award; the increase is .071 in the SFSP group, as shown in column 3 of Table 6, but this difference is not significantly different from zero at conventional levels. Paradoxically, males assigned to the SFP were less likely to meet the $1000 eligibility standard, though the negative effect is not quite significant. The estimates in column 2 show SFP males with an eligibility rate 6.6 percentage points lower than controls (s.e.=.037). There was no effect of the SFSP treatment on males' theoretical fellowship eligibility, while estimates for females indicate a large but only marginally significant gain in $5000 eligibility rates in column 9. The SSP does not appear to have affected fellowship eligibility.

Motivated by the findings which show larger effects on average first-year grades for students who had pre-registered for at least 4 courses, we also estimated the impact of treatment on fellowship eligibility in the four-course subsample. These results, reported in Panel B of

[17] Table 6 reports results for effects on eligibility status as determined by the program rules for GPA standards distributed to. In practice, payments were ultimately made using average 0-100 course grades instead of GPA. This results in a somewhat more generous award rate. Award eligibility was also lower than expected because overall grades worsened compared with previous years – the grade average for the control group, for example, was 3-4 percentage points lower than that in the previous two years.

Table 6, are generally similar to those for the full sample, reported in Panel A. However, both the positive SFSP effect on $5000 eligibility for females and the negative SFP effect on $1000 eligibility for males are larger in this sample.

In attempt to further understand the distribution shifts documented in Table 6, we also looked at the effects of SFP and SFSP on the entire GPA distribution. The results of this investigation, presented in a format similar to that used in Figure 2, appear in Figure 3. The cut-offs for fellowship eligibility are marked on the X-axes (the figure plots estimates using GPA's adjusted so that students from all high school grade quartiles have approximately the same cut-off). The SFSP effects on males are close to zero at every point in the distribution, as can be seen in Figure 3a. There are some significant negative SFP effects on males in the middle of the distribution, while the estimates for females show mostly positive effects, some significant. The GPA estimates for females are shown in Figure 3b. Some of the significant SFSP effects are in the neighbourhood of the award cut-offs but shifts in the grade distribution associated with the SFP are at lower levels.

B. Interaction-terms

We estimated treatment effects incorporating interactions with two variables, high school GPA and a binary indicator of students' planned hours of (paid) work, as reported in the background survey. The model with interaction terms is the same as equation (1), except that the second interaction term is a dummy to indicate students who plan to work more than 5 hours per week (the approximate median of this variable). These results are reported in Table 7, using a format similar to Table 3c. The GPA interaction (a dummy for the lowest quartile) is partly motivated by the possibility that students who did poorly in high school may be especially likely

to benefit from SSP-type services. On the other hand, males with low high school GPA made less use of services, so we might expect SSP effects to be larger in the high-GPA group. An interaction with students' (planned) work might arise if those who plan to hold down a part-time placed a higher value on the fellowships.

In practice, there is no clear pattern of treatment-effect variation with either the high school GPA or planned-work covariates. There are no significant interaction terms in the SSP estimates for males or females. Although some of the GPA interactions with the SFP effect are negative for males, the corresponding main effect is negative, so that the effect on low-GPA males comes out as essentially zero. All the SFSP interactions with low GPA are positive, but only one is (marginally) significant, that for females in the four-course sample. The total effects in this specification, reported in column 6, are also marginally significant. Given the number of coefficients and total effects in the table, this evidence does not seem strong enough to conclude that low-GPA females benefit more from fellowships.

C. Effects on Course Credit and Second-year Enrolment

Two primary goals of the STAR demonstration were to increase the number of credit-units students complete in freshman year, and to increase the likelihood that they go on to second year without dropping out of school. Estimates of the effect of the SSP, SFP, and SFSP on first-year credits and second-year retention rates are presented in Table 8. The average number of credits completed in the control group was 3.4, and 86 percent of controls registered for a second year at UTM. Estimates from the sample of all students with grades offer no evidence of substantial or significant differences in course completion, as can be seen in columns 1-3 of Panel A. Similarly, there is no evidence of an effect on credits earned in estimates using samples

limited to students who pre-registered for at least 4 or 5 courses. These estimates appear in Panels B and C.

In contrast with the effects on credits earned, the estimated effects on retention offer a hint of increased second-year enrolment in some samples and groups. Consistent with the estimated effects on achievement, there is some evidence for an increase in second-year retention for females in the SFSP group who pre-registered for at least four courses. The estimate for this group is .064 (s.e.=.03). The effect on females in the SFSP is large enough that, in combination with an insignificant positive effect on males, the overall SFSP effect on retention is significant in the samples with at least four and five courses (.054 with an s.e.=.026 in the former; .058 with an s.e.=.032 in the latter). Male retention rates were also higher in the SFP group, though this effect is less consistent across samples and does not generate a significant SFP effect when males and females are pooled. The full-sample SFP estimate for males is .064 (s.e.=.032), but this falls to .056 (s.e.=.035) in the four-course sample.

The significant positive effects reported in columns 1-3 constitute a substantial reduction in dropout rates. For example, the 6.4 percentage point estimate for SFSP females taking four or more courses constitutes a 7 percent increase in retention rates but almost a 50 percent decline in dropout rates. The corresponding magnitudes for the combined sample of boys and girls are 6 percent and 45 percent, relative to dropout rates of about 12 percent. It is worth emphasizing, however, that the effects for subgroups did not translate into much of an effect on overall retention rates. Moreover, the effect on males in the SFP is not supported by a corresponding impact on achievement. A group of potential male dropouts with fellowship-qualifying GPAs may have been encouraged to enrol in second year so as to collect fellowships (recall that second-year enrolment was a prerequisite for payment). On the other hand, the smaller estimate

for males in the four-course sample suggests the significant estimate for males in the full sample may also be a chance finding.

D. Two-Stage Least Squares Estimates

Intention-to-treat effects are diluted by the fact that some of the students who were offered one of the STAR treatments failed to sign a statement of informed consent. Students who did not consent were ineligible for services and/or fellowships. We therefore refer to those who consented as participants. Effects of treatment on participants are of policy interest because they provide a better indicator of the average program effect in a world where everyone participates. 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 consent rate. Estimates of effects on participants in models with covariates can be conveniently constructed by 2SLS (Imbens and Angrist, 1994).

Effects of treatment on participants were constructed using two models. The first parallels the reduced forms allowing separate effects for each program. In particular, we estimated:

$$y_i = X_i'\delta + \alpha ssp_i^* + \beta sfp_i^* + \gamma sfsp_i^* + \eta_i, \qquad (2)$$

where ssp_i^*, sfp_i^*, and $sfsp_i^*$ indicate program participants. The participation variables were treated as endogenous and the three program-assignment dummies (ssp_i, sfp_i, and $sfsp_i$) used as instruments. Motivated by the absence of SSP effects, a variation on this model sets $\beta=\gamma$, equivalent to assuming that program effects in the SFSP are driven entirely by the offer of fellowships. This leads to an over-identified model with three instruments for two endogenous variables,

$$y_i = X_i'\delta + \alpha ssp_i^* + \beta anysfp_i^* + \eta_i, \quad (3)$$

where $anysfp_i^* = sfp_i^* + sfsp_i^*$. The endogenous $anysfp_i^*$ variable indicates participants in either the SFP or SFSP since the two fellowship treatments were mutually exclusive. In both cases, the 2SLS estimates are reported for models including basic covariates only; other covariate sets generate similar results.

Not surprisingly, given the reduced-form (intention-to-treat) results discussed above, none of the 2SLS estimates of effects on first-year grades in the full sample are significant. This can be seen in Panel A of Table 9a. Nevertheless, the 2SLS standard errors provide information – not immediately apparent in the reduced-form – about the size of detectable effects. For example, the fact that the effect of $anysfp_i^*$ in column 4 is estimated with a standard error of .86 means that conventional hypothesis tests have the power to detect of about .13σ (=2*.86/12.9) in this specification. Sex- and program-specific estimates are naturally less precise, but the level of precision is still such that fellowship effects of a reasonable magnitude should be detectable. Lower consent rates for the SSP, however, increase the standard errors associated with 2SLS estimates of SSP effects beyond those for the SFP (though not beyond those for the SFSP, since the SFSP treatment involved a smaller sample).

When the sample is limited to students who pre-registered for at least four courses, the estimated effect of the SFSP on first-year grades of females is marginally significant and on the order of .28σ (3.5 points, s.e.=1.9; see Panel B of Table 9a) The effect on females in this case is large enough that the corresponding overall effect is also significant, at around .22σ (2.9 points; s.e.=1.5) The SFSP effects on females are even larger in the sample of students planning to take at least 5 courses, as can be seen in Panel C. The overall effect on males and females in this sample is close to the corresponding estimate in the four-course sample.

2SLS estimates of retention effects, reported in Table 9b, parallel the reduced-form estimates**.** In particular, the 2SLS estimates in Panel A show an SFP effect on male participants of .081 (s.e.=.04), and a marginally significant Any SFP effect of .075 (s.e.=.037). In the sample of male students who took at least four courses, however, the estimates are no longer significant. In contrast, the 2SLS estimate of the effect of the SFSP on female retention is .08 (s.e.=.037) in the four-course sample, while the corresponding Any SFP effect on females is .064 (s.e.=.037). The SFSP effects on females are large enough to generate a significant overall retention effect in this sample of .071 (s.e.=.033). Estimates in the sample of students who took at least five course are similar, though somewhat larger for SFP males and SFSP females. The overall effect of the SFSP on retention in this sample is .079 (s.e.=.044), again a result driven by females.

VI. Student Reports

In order to better understand students' perception of the program and their reaction to it, we conducted a few open-ended interviews with students in each of the treatment groups, sampling from those who consented.[18] 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, and could react to statements by others in the group. Interviews lasted about one hour and were guided by a list of questions.

Given the generally modest impact of the offer of fellowships, a key concern is whether program participants were aware of and understood the fellowship program. The five SFP participants we interviewed indicated they were aware of the program. Moreover, all seem to

[18] Oreopoulos interviewed one female student who did not consent to participate in the SFP. Asked why she did not consent, the student responded, "Honestly I can't remember why I didn't pursue it. I don't know why. As a first year, I was really nervous starting and overwhelmed. Now that I think of it, I probably should have joined".

have understood the basic idea, though a few wondered at first if "it was too good to be true." The two SFSP participants we interviewed also indicated an initial scepticism about the $5000 scholarship, but signed up to participate within the first week of the offer. Most of the SFP participants found the reminder cards helpful; one indicated they kept one in their wallet. The students also remembered receiving email reminders from the STAR office and seemed to find the program communications effective.

Students generally saw the program as a motivator. One student reported that "I found it had a positive influence. I found my study habits improved. It might have been the money motivation. I would definitely say that it had a positive influence." In fact, four out of five SFP participants interviewed saw the program as beneficial; the other indicated, "It certainly did not detract from anything." On the other hand, some students reported they were not routinely mindful of the program or forgot about it completely.

Most of the students seem comfortable with the idea of cash awards. One SFSP 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." An SFP participant commented, "I wanted it. $5000! I definitely wasn't going to miss that." (This student did in fact receive the full scholarship.) 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 one that I enjoy and leaving the others in the dust." One student become discouraged, however, and indicated that once she realized she could not meet her target, the fellowship no longer provided motivation.

We talked to a total of 5 students who had access to services, two in the SFSP and three in the SSP. Here too, we were especially interested in whether students understood the nature of the services available to them. This seems to have been the case. On the other hand, a number noted major scheduling conflicts that prevented them from attending FSGs. There also seems to have been some peer interaction that affected attendance, with student decisions to attend influenced by whether their friends were in the program and could also attend. Our assignment mechanism did not exploit this as there was no deliberate clustering in the research design.

Eleven other students who were offered the SFP and SFSP were selected at random for one-on-one interviews and offered $10 gift certificates for 30 minute meetings. Five meetings occurred, all with females.[19] Reactions to the fellowships were similar to those interviewed in the focus groups, but some students that finished with particularly low grades expressed less or decreasing interest over time. One SFSP student who finished with a 1.2 GPA commented, "At first I was excited about it, 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."[20]

On balance, these discussions suggest the fellowship program penetrated students' consciousness, but in many cases interest declined. Surprisingly, the students we spoke with did not seem to see the fellowship targets as unrealistic, though clearly that was the case for most of them. Rather, the impact of the program seems to have faded as the year wore on and other

[19] The probability of showing up for these interviews appears unrelated to student GPAs.

[20] 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."

concerns became more pressing. In retrospect, the power of the fellowship as a motivator might have been boosted by more frequent contacts and by setting lower grade targets, though this would have increased costs. An important though logistically unavoidable issue with the FSG component was a number of scheduling conflicts, although other students said they did not go because they felt they had no time.

VII. Summary and Conclusions

Many college students perform poorly, drop courses, or drop out entirely. The goal of the Student Achievement and Retention (STAR) project was to learn more about the potential for support services and financial incentives to increase retention, and to foster academic skill and motivation. Incoming college freshman in randomly selected treatment groups were offered peer-advising and supplemental instruction services, merit awards, or both services and awards. The incentives came in the form of $1,000 cash payments for students who cleared GPA thresholds ranging from C+ to B, depending on high school GPA. Larger payments of $2,500 and $5,000 were offered to those clearing higher GPA targets ranging from B- to A-. To the best of our knowledge, STAR is the first large-scale randomized evaluation of incentives of this kind for college students.

Student interest in support services was lower than expected, and the students with the weakest academic background were least likely to give informed consent or to make use of services. On the other hand, interest in services as reflected in consent rates and service usage were both markedly higher in the group that was also offered cash incentives in the form of SFP fellowships. Interest and use were also much higher for females than males. Peer advising was considerably more popular than supplemental instruction for both sexes. The advising

intervention clearly bears further exploration, as does the use of achievement incentives to boost interest in services.

Results from the impact evaluation of STAR are not entirely clear-cut, but a number of patterns emerge. First, students offered services without fellowships did no better than those in the control group. This may be because take-up rates were low, but a 2SLS analysis that adjusts for take-up reveals a level of precision sufficient to detect service effects equal to about .25σ in the combined sample of males and females. Retention rates were also unaffected by services.

Second, the fellowship treatments had little overall impact on freshman-year GPAs. A closer look, however, reveals substantial effects on female students' grades at the end of the Fall term and somewhat more modest effects on the full-year grades for females taking close to a full course load. In addition to being of substantive interest, the strong results for the effect of fellowships treatments on the Fall term grades of females serves to establish the statistical power of the STAR research design. There are no significant positive effects on grades for males.

A related finding is the generally larger effect of the combined (SFSP) intervention relative to the offer of fellowships alone through the SFP. In addition to stronger achievement effects in this group, SFSP females taking at least four courses also appear to have increased retention. Two reasons for the SFP/SFSP difference seem likely. On one hand, students who were trying to win fellowships had access to services that may have helped them achieve their targets. The fact that service use was higher in the SFSP than for services alone is also consistent with this. In addition, students in the combined treatment group received more frequent contacts in the form of biweekly reminders from Peer Advisors.

We are left with the question of why the large effects on Fall term grades were not sustained. The explanation seems to have to do both with samples and timing. The Fall term

grades sample was more likely to be carrying something close to a full course load, and a near-full load was a pre-requisite for fellowships. The timing consideration has to do with the difficulty students encountered in boosting their performance throughout the year. Effects on achievement may also have faded due to difficulties students encountered in their attempt to meet award targets. For many students, these targets were out of reach, a fact that was likely to have become apparent at the end of Fall term. Other students simply forgot about the program or became distracted by other concerns. One lesson here may be that incentives are more likely to be effective when the award standards are easier and when the awards come more quickly, perhaps at the end of every term. In future evaluations of this sort, it may also be worth looking at payoff functions that are linear in achievement, as suggested by Holmstrom and Milgrom (1987).

Another interesting question is why females responded so much more than males, both in terms of take-up and impact. Although we have no simple explanation, it is worth noting that similar sex differences in program impact have been observed elsewhere. Dynarski (2005) estimates larger effects of tuition aid on college completion for women (in US states), while Girabaldi, *et al* (2006) find that tuition affects the completion rates of women more than men (in Italy). In a study of the effects of merit awards on high school students, Angrist and Lavy (2002) find effects on girls only (in Israel). Somewhat 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.

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 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. (2006), "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 (2004), "How Financial Aid Affects Persistence," NBER Working Paper no. 10242, January.

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

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.

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.

Garilbaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Rettore, "College Cost and Students' Performance: Evidence from Tuition Discontinuities," mimeo, 2006.

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.

Holmstrom, Bengt and Paul Milgrom (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55(2), pp. 303-28.

Jaeger, David and Marianne Page (1996), "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics*, 78(November), pp. 733-739.

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

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

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).

Loury, Linda Datcher and David Garman (1995), "College Selectivity and Earnings," *Journal of Labor Economics* 13(April), pp. 289-308.

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

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), pp. 2-9.

Oreopoulos, Philip (2006) "Do Dropouts Drop Out Too Soon? Wealth, Health, and Happiness from Compulsory Schooling," Mimeo, University of Toronto, Department of Economics.

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

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 – 25th percentile	2.3 (C+)	2.7 (B-)	3.0 (B)
25 – 50th percentile	2.7 (B-)	3.0 (B)	3.3 (B+)
50th – 75th 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.

Table 1: Descriptive Statistics

		Contrasts by treatment status				Obs	Fall Grades Sample	
	Control Mean	SSP v. Control	SFP v. Control	SFSP v. Control	F-stat (all=control)		Control Mean	Obs
Administrative variables								
Registered	0.965	0.019 [0.012]	0.019 [0.012]	-0.005 [0.015]	1.58 (0.193)	1656	1.00	1397
Took survey	0.989	0.011 [0.008]	-0.009 [0.008]	-0.009 [0.009]	1.83 (0.139)		0.992	
Completed survey	0.888	-0.020 [0.023]	-0.012 [0.023]	-0.054 [0.029]	1.31 (0.271)		0.907	
Canada to analyze academic and income data	0.762	-0.014 [0.030]	-0.030 [0.030]	-0.002 [0.038]	0.37 (0.777)		0.780	
Has fall grades	0.844	-0.004 [0.026]	0.032 [0.026]	-0.051 [0.032]	1.63 (0.181)		1.00	
Courses enrolled as of Fall 2005	4.05 {1.38}	-0.027 [0.098]	0.013 [0.088]	-0.184 [0.132]	0.70 (0.550)		4.34 {1.01}	
Student background variables								
Female	0.571	-0.003 [0.035]	0.029 [0.035]	-0.024 [0.043]	0.41 (0.749)		0.563	
High school GPA	78.7 {4.23}	0.175 [0.301]	0.148 [0.301]	-0.197 [0.373]	0.32 (0.812)		78.8 {4.27}	
Age	18.3 {0.628}	-0.012 [0.045]	-0.020 [0.045]	0.041 [0.055]	0.33 (0.805)		18.3 {0.600}	
Survey response variables								
Hrs/wk study in high school	17.7 {12.7}	-0.644 [0.921]	-0.425 [0.917]	-0.492 [1.162]	0.23 (0.879)	1454	17.9 {12.6}	1273
Attending first choice university	0.244	0.009 [0.033]	0.062 [0.033]	0.036 [0.042]	1.29 (0.277)		0.232	
Parents very important in uni. decision	0.400	-0.008 [0.037]	-0.034 [0.037]	-0.024 [0.047]	0.34 (0.798)		0.407	
Sure about career	0.363	0.038 [0.036]	0.016 [0.036]	-0.059 [0.046]	1.14 (0.333)		0.357	
Concerned about funding	0.386	0.028 [0.037]	0.034 [0.037]	-0.034 [0.047]	0.72 (0.542)		0.395	
Plans to work at university (any)	0.781	0.031 [0.031]	-0.068 [0.031]	0.035 [0.040]	2.65 (0.048)		0.773	
Hrs/wk planned study at university	27.4 {14.3}	1.40 [1.10]	-0.488 [1.10]	-1.03 [1.39]	0.96 (0.409)		27.6 {14.2}	
Often procrastinate	0.354	-0.036 [0.036]	0.002 [0.036]	-0.082 [0.045]	1.34 (0.258)		0.366	
≥B grade avg very important	0.849	0.022 [0.026]	0.023 [0.026]	-0.025 [0.036]	0.72 (0.541)		0.852	

Table 1: Descriptive Statistics, continued

		Contrasts by treatment status				Obs	Fall Grades Sample	
	Control Mean	SSP v. Control	SFP v. Control	SFSP v. Control	F-stat (all=control)		Control Mean	Obs
Survey response variables, cont.								
Commutes from home	0.803	-0.033 [0.032]	0.019 [0.029]	0.005 [0.038]	0.64 (0.592)	1454	0.805	1273
Plans to do graduate studies	0.553	0.060 [0.037]	-0.028 [0.038]	0.079 [0.046]	2.13 (0.094)		0.555	
Plans to finish in 4 years	0.816	-0.010 [0.030]	0.006 [0.029]	-0.048 [0.040]	0.55 (0.647)		0.836	
Family background variables								
Mother tongue is English	0.701	0.019 [0.032]	0.007 [0.032]	0.046 [0.040]	0.50 (0.681)	1656	0.688	1397
Mother tongue is French	0.005	-0.001 [0.005]	-0.001 [0.005]	-0.005 [0.006]	0.26 (0.856)		0.006	
Mom graduated from high school	0.869	0.016 [0.026]	-0.024 [0.026]	-0.045 [0.033]	1.12 (0.338)	1454	0.863	1273
Mom graduated from college	0.364	0.042 [0.036]	-0.031 [0.036]	-0.060 [0.046]	1.46 (0.223)		0.361	
Dad graduated from high school	0.839	0.023 [0.028]	0.006 [0.028]	-0.015 [0.035]	0.34 (0.796)		0.84	
Dad graduated from college	0.452	0.013 [0.038]	-0.009 [0.038]	-0.044 [0.048]	0.39 (0.763)		0.465	

Notes: Mean standard deviations in curly brackets "{}". Regression standard errors in straight brackets "[ ]". F-test p-values in parentheses.

Table 2a: Selection Effects

	Registered			Has grades			Has Fall Grades		
	No controls	Basic controls	All controls	No controls	Basic controls	All controls	No controls	Basic controls	All controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control group mean		0.965 (0.183)			0.942 (0.233)			0.844 (0.363)	
Offered SSP	0.019	0.019	0.008	-0.006	-0.004	0.006	-0.004	-0.002	0.001
	[0.010]	[0.009]	[0.008]	[0.017]	[0.011]	[0.010]	[0.026]	[0.022]	[0.022]
Offered SFP	0.019	0.012	0.017	0.026	0.014	0.02	0.032	0.028	0.021
	[0.010]	[0.008]	[0.005]	[0.013]	[0.011]	[0.010]	[0.024]	[0.022]	[0.024]
Offered SSP and SFP	-0.005	0.003	-0.001	-0.029	-0.013	-0.007	-0.051	-0.019	-0.009
	[0.017]	[0.014]	[0.012]	[0.024]	[0.017]	[0.015]	[0.035]	[0.031]	[0.033]
Observations	1656	1656	1461	1656	1656	1461	1656	1656	1461

Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control group mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below.

Sample in columns (1), (2), (4), (5), (7) and (8) is all students participating in the STAR program. Sample in columns (3), (6) and (9) is all STAR students who completed an online questionnaire. Sample in columns (10), (11), (13), (14), (16) and (17) is all STAR students matched to UTM grades data as of June, 2006. Sample in columns (12), (15) and (18) is students matched to both grades and questionnaire data. Basic controls include sex, mother tongue, high school grade quartile and number of credits enrolled. All controls add responses to 5 survey questions: Was UTM your first-choice university, How many hours/week did you study in high school,How many hours/week do you plan to work while in school, What are your mother's and father's education levels.

Table 2b: Selection Effects (continued)

	Number of credits completed			Number of fall credits completed			Number of math and science credits completed		
	No controls (10)	Basic controls (11)	All controls (12)	No controls (13)	Basic controls (14)	All controls (15)	No controls (16)	Basic controls (17)	All controls (18)
Control group mean		3.95 (0.921)			0.74 (0.446)			1.10 (1.21)	
Offered SSP	0.061 [0.064]	0.067 [0.047]	0.041 [0.051]	0.029 [0.032]	0.031 [0.031]	0.022 [0.032]	0.080 [0.091]	0.08 [0.084]	0.111 [0.088]
Offered SFP	-0.011 [0.065]	0.016 [0.052]	0.02 [0.053]	-0.052 [0.029]	-0.033 [0.028]	-0.036 [0.030]	0.103 [0.086]	0.097 [0.081]	0.113 [0.085]
Offered SSP and SFP	-0.088 [0.084]	0.004 [0.060]	0.006 [0.064]	0.009 [0.044]	0.023 [0.041]	0.028 [0.044]	-0.180 [0.104]	-0.108 [0.093]	-0.080 [0.102]
Observations	1561	1561	1410	1561	1561	1410	1561	1561	1410

Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control group mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below.

Sample in columns (1), (2), (4), (5), (7) and (8) is all students participating in the STAR program. Sample in columns (3), (6) and (9) is all STAR students who completed an online questionnaire. Sample in columns (10), (11), (13), (14), (16) and (17) is all STAR students matched to UTM grades data as of June, 2006. Sample in columns (12), (15) and (18) is students matched to both grades and questionnaire data. Basic controls include sex, mother tongue, high school grade quartile and number of credits enrolled. All controls add responses to 5 survey questions: Was UTM your first-choice university, How many hours/week did you study in high school,How many hours/week do you plan to work while in school, What are your mother's and father's education levels.

Table 3a: First-stage Effects

	Responded to STAR Invitation (STAR Participant)		Used SSP Services, Fall		Used SSP Services, Spring	
	Basic controls (1)	All controls (2)	Basic controls (3)	All controls (4)	Basic controls (5)	All controls (6)
			Panel A: All students			
Control group mean	-0-	-0-	-0-	-0-	-0-	-0-
Offered SSP	0.504	0.551	0.160	0.178	0.156	0.169
	[0.032]***	[0.034]***	[0.023]	[0.026]	[0.023]	[0.025]
Offered SFP	0.854	0.867	-0.003	-0.004	-0.001	-0.001
	[0.022]***	[0.022]***	[0.002]	[0.003]	[0.002]	[0.003]
Offered SSP and SFP	0.738	0.783	0.282	0.288	0.254	0.284
	[0.036]***	[0.037]***	[0.037]	[0.041]	[0.036]	[0.041]
Observations	1607	1429	1607	1429	1607	1429
			Panel B: Males			
Control group mean	-0-	-0-	-0-	-0-	-0-	-0-
Offered SSP	0.431	0.459	0.083	0.091	0.133	0.141
	[0.047]***	[0.051]***	[0.027]	[0.030]	[0.033]	[0.036]
Offered SFP	0.789	0.814	0.000	0.002	0.002	0.006
	[0.041]***	[0.040]***	[0.003]	[0.005]	[0.003]	[0.005]
Offered SSP and SFP	0.669	0.708	0.192	0.200	0.193	0.222
	[0.057]***	[0.063]***	[0.049]	[0.056]	[0.050]	[0.058]
Observations	683	602	683	602	683	602
			Panel C: Females			
Control group mean	-0-	-0-	-0-	-0-	-0-	-0-
Offered SSP	0.559	0.623	0.218	0.242	0.173	0.192
	[0.042]***	[0.044]***	[0.035]	[0.039]	[0.032]	[0.036]
Offered SFP	0.903	0.908	-0.004	-0.007	-0.002	-0.004
	[0.024]***	[0.025]***	[0.003]	[0.005]	[0.002]	[0.005]
Offered SSP and SFP	0.790	0.837	0.354	0.355	0.304	0.332
	[0.046]***	[0.044]***	[0.053]	[0.057]	[0.051]	[0.056]
Observations	924	827	924	827	924	827

Notes: Heteroskedasticity-robust standard errors in brackets. Sample in columns (1), (3) and (5) is all registered students participating in the STAR program. Sample in remaining columns is registered STAR students who completed an online questionnaire. Basic controls include sex, mother tongue, high school grade quartile and number of courses enrolled. All controls add responses to 5 survey questions: Is this your first-choice university, How many hours/week did you study in high school, How many hours/week do you plan to work while in school, What are your mother' and father's education levels.

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

Table 3b: SSP Take-up by Service and Group

First-stage variable	Received SSP Services		Met with/emailed an Advisor		Attended FSGs	
	Full sample (1)	Fall grades sample (2)	Full sample (3)	Fall grades sample (4)	Full sample (5)	Fall grades sample (6)
			Panel A: All			
Offered SSP	0.228	0.253	0.195	0.215	0.102	0.119
	[0.027]	[0.030]	[0.025]	[0.028]	[0.019]	[0.022]
Offered SSP and SFP	0.390	0.395	0.361	0.378	0.126	0.118
	[0.040]	[0.045]	[0.040]	[0.044]	[0.028]	[0.030]
Observations	1607	1397	1607	1397	1607	1397
			Panel B: Males			
Offered SSP	0.190	0.211	0.143	0.159	0.094	0.104
	[0.038]	[0.042]	[0.034]	[0.038]	[0.028]	[0.031]
Offered SSP and SFP	0.264	0.272	0.248	0.273	0.109	0.096
	[0.055]	[0.063]	[0.054]	[0.062]	[0.040]	[0.042]
Observations	683	602	683	602	683	602
			Panel C: Females			
Offered SSP	0.257	0.287	0.236	0.261	0.107	0.130
	[0.037]	[0.042]	[0.036]	[0.041]	[0.026]	[0.031]
Offered SSP and SFP	0.489	0.488	0.450	0.457	0.140	0.136
	[0.056]	[0.061]	[0.056]	[0.061]	[0.039]	[0.041]
Observations	924	795	924	795	924	795

Notes: Heteroscedasticity-robust standard errors in brackets. Sample in columns (1), (3) and (5) is all enrolled students participating in the STAR program with at least one grade as of June, 2006. Sample in columns (2), (5) and (8) is enrolled STAR students with at least one fall grade. All regressions control for mother tongue, and high school grade quartile. Panel A also

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

Table 3c: Interactions in Service take-up effects

	STAR Consent			Used SSP Services		
	All	Males	Females	All	Males	Females
Program	(1)	(2)	(3)	(4)	(5)	(6)
SSP	0.405 [0.083]***	0.422 [0.122]***	0.420 [0.116]***	0.170 [0.058]***	0.152 [0.077]**	0.190 [0.087]**
× low HS GPA	-0.118 [0.063]*	-0.224 [0.094]**	-0.042 [0.085]	-0.129 [0.053]**	-0.171 [0.073]**	-0.102 [0.075]
Total effect	0.287 [0.080]	0.198 [0.116]	0.378 [0.118]	0.041 [0.053]	-0.019 [0.061]	0.088 [0.090]
× ≥4 courses	0.186 [0.083]**	0.140 [0.121]	0.187 [0.119]	0.146 [0.057]**	0.142 [0.072]*	0.142 [0.091]
Total effect	0.473 [0.047]	0.339 [0.071]	0.565 [0.061]	0.187 [0.036]	0.123 [0.048]	0.230 [0.051]
SFP	0.687 [0.084]***	0.533 [0.127]***	0.867 [0.092]***	—	—	—
× low HS GPA	-0.043 [0.044]	-0.039 [0.080]	-0.031 [0.050]	—	—	—
Total effect	0.645 [0.085]	0.494 [0.124]	0.836 [0.094]			
× ≥4 courses	0.218 [0.085]**	0.341 [0.125]***	0.056 [0.093]	—	—	—
Total effect	0.862 [0.035]	0.834 [0.059]	0.891 [0.040]			
SFSP	0.619 [0.094]***	0.516 [0.135]***	0.741 [0.124]***	0.248 [0.100]**	0.266 [0.124]**	0.260 [0.168]
× low HS GPA	-0.030 [0.073]	-0.072 [0.117]	-0.012 [0.093]	0.062 [0.082]	-0.029 [0.111]	0.099 [0.119]
Total effect	0.590 [0.090]	0.445 [0.131]	0.729 [0.121]	0.310 [0.084]	0.237 [0.116]	0.360 [0.128]
× ≥4 courses	0.164 [0.093]*	0.243 [0.137]*	0.066 [0.124]	0.141 [0.095]	0.026 [0.126]	0.203 [0.147]
Total effect	0.754 [0.053]	0.688 [0.098]	0.795 [0.060]	0.451 [0.062]	0.263 [0.093]	0.563 [0.077]
Observations	1607	683	924	1607	683	924

Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Rows labeled total effect report the point estimate for the group effect plus the interaction, followed by its standard error in brackets. Sample is all registered students participating in the STAR program. All regressions control for high school grade quartile, mother tongue, and number of courses. Columns (1) and (4) also control for sex.

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

Table 4: Treatment Effect on Fall and First-year Grade (Fall grades sample)

	Fall Grade			First-year Grade		
Program	No Controls (1)	Basic (2)	All (3)	No Controls (4)	Basic (5)	All (6)
			Panel A: All Students with Fall Grades			
Control mean		63.8 (12.1)			60.2 (12.9)	
SSP	0.370 [0.964]	0.211 [0.896]	0.382 [0.916]	0.353 [1.058]	0.159 [0.957]	0.734 [0.946]
SFP	2.01 [0.860]**	1.88 [0.844]**	1.77 [0.867]**	0.999 [0.847]	0.910 [0.812]	0.399 [0.869]
SFSP	1.75 [1.190]	1.89 [1.170]	2.59 [1.261]**	2.08 [1.221]*	2.30 [1.168]**	2.74 [1.269]**
Observations	1397	1397	1263	1397	1397	1263
			Panel B: Males			
Control mean		65.1 (11.9)			61.0 (13.4)	
SSP	-0.406 [1.512]	-0.607 [1.490]	-0.162 [1.458]	-0.967 [1.743]	-1.086 [1.669]	-0.613 [1.625]
SFP	0.817 [1.286]	0.743 [1.249]	1.31 [1.226]	-0.351 [1.306]	-0.142 [1.302]	-0.592 [1.407]
SFSP	-0.196 [1.771]	-0.208 [1.718]	1.07 [1.916]	1.39 [1.952]	1.08 [1.913]	1.80 [2.174]
Observations	602	602	538	602	602	538
			Panel C: Females			
Control mean		62.8 (12.1)			59.6 (12.4)	
SSP	0.944 [1.240]	1.25 [1.066]	1.55 [1.150]	1.40 [1.293]	1.44 [1.075]	2.39 [1.072]**
SFP	2.96 [1.147]**	2.74 [1.128]**	2.46 [1.237]**	1.97 [1.109]*	1.73 [1.041]*	1.51 [1.145]
SFSP	3.24 [1.604]**	3.53 [1.557]**	3.84 [1.653]**	2.61 [1.560]*	3.33 [1.477]**	3.72 [1.576]**
Observations	795	795	725	795	795	725

Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Sample is all enrolled students participating in the STAR program with at least one fall grade as of May, 2006; cols (3) and (6) is those with a fall grade who completed an online questionnaire. Basic controls include high school grade quartile, mother tongue, high school grade and number of courses. All controls add responses to 5 survey questions: Is this your first-choice university, How many hours/week did you study in high school,How many hours/week do you plan to work while in school, What are your mother's and father's education levels. Panel A "Basic" and "All" regressions also control for sex.

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

Table 5: Treatment Effect on First-year Grade (Credit-weighted)

	By type			Pooled		
	All	Males	Females	All	Males	Females
Program	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: All students			
Control mean	60.1 (12.9)	60.8 (13.2)	59.7 (12.7)	60.1 (12.9)	60.8 (13.2)	59.7 (12.7)
SSP	-0.832 [0.917]	-1.24 [1.590]	-0.249 [1.075]	-0.832 [0.917]	-1.24 [1.588]	-0.248 [1.074]
SFP (Any)				0.644 [0.718]	-0.396 [1.138]	1.37 [0.926]
SFP	0.249 [0.801]	-0.774 [1.310]	1.01 [1.011]			
SFSP	1.34 [1.175]	0.216 [1.778]	2.06 [1.576]			
Observations	1561	661	900	1561	661	900
			Panel B: Students with Fall Grades			
Control mean	60.2 (12.9)	61.0 (13.4)	59.6 (12.4)	60.2 (12.9)	61.0 (13.4)	59.6 (12.4)
SSP	0.159 [0.957]	-1.09 [1.669]	1.44 [1.075]	0.161 [0.957]	-1.08 [1.667]	1.44 [1.074]
SFP (Any)				1.40 [0.723]*	0.313 [1.170]	2.27 [0.919]**
SFP	0.910 [0.812]	-0.142 [1.302]	1.73 [1.041]*			
SFSP	2.30 [1.168]**	1.08 [1.913]	3.33 [1.477]**			
Observations	1397	602	795	1397	602	795

(cont.)

Table 5: Treatment Effect on First-year Grade, cont.

	By type			Pooled		
	All	Males	Females	All	Males	Females
Program	(1)	(2)	(3)	(4)	(5)	(6)
			Panel C: Students with at least 5 courses			
Control mean	62.1 (11.2)	63.4 (10.8)	61.2 (11.3)	62.1 (11.2)	63.4 (10.8)	61.2 (11.3)
SSP	0.343 [1.204]	-1.85 [2.194]	1.75 [1.339]	0.343 [1.203]	-1.85 [2.190]	1.74 [1.338]
SFP (Any)				0.867 [0.952]	-1.10 [1.603]	1.96 [1.189]*
SFP	-0.007 [1.031]	-1.214 [1.486]	0.622 [1.353]			
SFSP	2.29 [1.631]	-0.943 [3.004]	4.41 [1.878]**			
Observations	694	264	430	694	264	430
			Panel D: Students with at least 4 courses			
Control mean	60.8 (12.5)	61.4 (12.9)	60.3 (12.2)	60.8 (12.5)	61.4 (12.9)	60.3 (12.2)
SSP	-0.218 [0.919]	-0.332 [1.535]	-0.122 [1.128]	-0.217 [0.918]	-0.318 [1.534]	-0.122 [1.127]
SFP (Any)				0.522 [0.741]	-0.743 [1.211]	1.39 [0.938]
SFP	-0.393 [0.850]	-2.11 [1.419]	0.692 [1.060]			
SFSP	2.26 [1.159]*	1.57 [1.808]	2.82 [1.535]*			
Observations	1345	554	791	1345	554	791

Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Sample is all enrolled students participating in the STAR program with at least one grade as of May, 2006, restricted as noted in each panel. All regressions control for high school grade quartile, number of credit dummies and mother tongue. Columns (1) and (4) also control for sex. This specification corresponds to "Basic" controls in Table 5.

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

Table 6: Treatment Effect on Fellowship Eligibility

	Pooled								
	$1,000+			$2,500			$5,000		
	All	Males	Females	All	Males	Females	All	Males	Females
Program	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A: Students with Grades					
Control	0.161	0.191	0.140	0.084	0.097	0.075	0.033	0.040	0.028
mean	0.368	0.393	0.347	0.278	0.296	0.264	0.178	0.195	0.164
SSP	-0.013	-0.048	0.010	-0.021	-0.040	-0.009	-0.008	-0.004	-0.013
	[0.025]	[0.038]	[0.034]	[0.018]	[0.027]	[0.024]	[0.012]	[0.022]	[0.013]
SFP	-0.011	-0.066	0.030	-0.028	-0.020	-0.030	-0.004	-0.014	0.005
	[0.025]	[0.037]	[0.034]	[0.017]	[0.029]	[0.022]	[0.012]	[0.018]	[0.017]
SFSP	0.040	0.003	0.071	0.015	-0.012	0.037	0.044	0.017	0.067
	[0.035]	[0.053]	[0.047]	[0.026]	[0.037]	[0.036]	[0.023]	[0.029]	[0.034]*
Observations	1561	661	900	1561	661	900	1561	661	900
				Panel B: Students with ≥4 Courses					
Control	0.182	0.222	0.154	0.094	0.111	0.082	0.037	0.047	0.029
mean	(0.386)	(0.416)	(0.361)	(0.292)	(0.315)	(0.275)	(0.188)	(0.211)	(0.169)
SSP	-0.009	-0.055	0.020	-0.020	-0.044	-0.003	-0.007	-0.006	-0.012
	[0.028]	[0.043]	[0.037]	[0.020]	[0.031]	[0.027]	[0.013]	[0.025]	[0.015]
SFP	-0.018	-0.084	0.025	-0.026	-0.018	-0.028	-0.002	-0.016	0.008
	[0.028]	[0.043]	[0.037]	[0.020]	[0.035]	[0.024]	[0.014]	[0.022]	[0.018]
SFSP	0.038	-0.006	0.073	0.025	-0.007	0.048	0.055	0.024	0.080
	[0.041]	[0.065]	[0.054]	[0.031]	[0.047]	[0.042]	[0.028]*	[0.037]	[0.039]*
Observations	1345	554	791	1345	554	791	1345	554	791

Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Sample in Panel A is all enrolled students participating in the STAR program with at least one grade as of May, 2006. Sample in Panel B is all STAR students enrolled in 4 or more courses. All regressions control for high school grade quartile, number of courses and mother tongue. Columns (1), (4) and (7) also control for sex.

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

Table 7: Interaction Effects for First-year Grade

	All Students			Students with ≥4 Courses		
	All	Males	Females	All	Males	Females
Program	(1)	(2)	(3)	(4)	(5)	(6)
Control mean	60.1 (12.9)	60.8 (13.2)	59.7 (12.7)	60.8 (12.5)	61.4 (12.9)	60.3 (12.2)
SSP	-0.241 [1.701]	-1.27 [2.737]	0.548 [2.219]	0.209 [1.736]	-1.52 [2.909]	1.43 [2.198]
× low HS GPA	0.023 [1.764]	3.85 [3.054]	-2.64 [2.069]	-0.102 [1.768]	2.29 [3.076]	-1.67 [2.131]
Total effect	-0.218 [1.756]	2.579 [2.970]	-2.091 [2.187]	0.106 [1.746]	0.768 [3.076]	-0.240 [2.153]
× work≥5 hrs/wk	-0.088 [1.834]	-2.06 [3.063]	1.66 [2.254]	0.425 [1.818]	0.593 [3.137]	0.339 [2.261]
Total effect	-0.306 [1.473]	0.516 [2.524]	-0.435 [1.771]	0.531 [1.567]	1.362 [2.796]	0.099 [1.845]
SFP	-2.71 [1.602]*	-8.37 [2.873]***	-0.079 [1.876]	-3.77 [1.711]**	-9.62 [3.144]***	-1.01 [1.991]
× low HS GPA	3.22 [1.713]*	7.40 [2.749]***	1.62 [2.202]	3.31 [1.817]*	6.20 [3.069]**	2.81 [2.290]
Total effect	0.512 [1.963]	-0.974 [3.465]	1.543 [2.381]	-0.457 [2.092]	-3.427 [3.942]	1.801 [2.483]
× work≥5 hrs/wk	1.41 [1.838]	4.43 [3.274]	0.003 [2.202]	2.35 [1.963]	5.78 [3.653]	0.474 [2.325]
Total effect	1.918 [1.391]	3.456 [1.998]	1.547 [1.899]	1.889 [1.480]	2.355 [2.170]	2.275 [2.009]
SFSP	0.968 [2.555]	-1.35 [4.150]	3.17 [3.273]	0.108 [2.678]	-4.60 [4.708]	2.89 [3.226]
× low HS GPA	3.29 [2.530]	2.14 [4.140]	4.29 [3.267]	3.94 [2.469]	3.67 [4.018]	5.06 [3.033]*
Total effect	4.253 [2.886]	0.793 [4.963]	7.466 [3.084]	4.051 [3.059]	-0.923 [5.433]	7.950 [3.009]
× work≥5 hrs/wk	-1.34 [2.837]	1.03 [4.684]	-4.02 [3.306]	0.595 [2.945]	5.66 [5.054]	-3.11 [3.234]
Total effect	2.917 [2.055]	1.826 [3.645]	3.443 [2.463]	4.646 [2.087]	4.738 [3.512]	4.841 [2.500]
Observations	1421	590	831	1242	507	735

Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Rows labeled total effect report the point estimate for the group effect plus the interaction, followed by its standard error in brackets. Sample is all registered students participating in the STAR program who completed an online questionnaire. All regressions control for high school grade quartile, mother tongue, and number of courses. Columns (1) and (4) also control for sex

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

Table 8: Treatment Effect on First-year Credits and Second-year Enrollment

	First-year credits earned			Enrolled for second-year (yes=1)		
Program	All (1)	Males (2)	Females (3)	All (4)	Males (5)	Females (6)
			Panel A: All Students			
Control mean	3.35 (1.52)	3.27 (1.50)	3.41 (1.52)	0.858 (0.349)	0.855 (0.352)	0.860 (0.348)
SSP	0.015 [0.089]	-0.082 [0.135]	0.070 [0.119]	-0.022 [0.025]	0.016 [0.035]	-0.058 [0.035]
SFP	0.081 [0.088]	-0.007 [0.140]	0.146 [0.114]	0.030 [0.023]	0.064 [0.032]**	0.004 [0.032]
SFSP	-0.035 [0.109]	-0.125 [0.172]	0.052 [0.141]	0.037 [0.027]	0.040 [0.041]	0.036 [0.035]
Observations	1607	683	924	1607	683	924
			Panel B: Students with ≥4 Courses			
Control mean	3.61 (1.40)	3.54 (1.39)	3.66 (1.40)	0.883 (0.322)	0.874 (0.332)	0.874 (0.332)
SSP	0.065 [0.092]	0.003 [0.139]	0.085 [0.124]	0.001 [0.025]	0.042 [0.034]	-0.033 [0.035]
SFP	-0.011 [0.090]	-0.091 [0.149]	0.046 [0.113]	0.013 [0.024]	0.056 [0.035]	-0.012 [0.032]
SFSP	0.066 [0.117]	0.057 [0.193]	0.101 [0.145]	0.054 [0.026]**	0.044 [0.044]	0.064 [0.030]**
Observations	1352	557	795	1352	557	795
			Panel C: 5+ courses sample			
Control mean	4.08 (1.36)	4.08 (1.36)	4.08 (1.36)	0.890 (0.313)	0.883 (0.322)	0.894 (0.308)
SSP	0.181 [0.129]	0.058 [0.211]	0.239 [0.161]	0.031 [0.030]	0.040 [0.047]	0.024 [0.039]
SFP	0.100 [0.124]	0.025 [0.212]	0.141 [0.157]	0.028 [0.032]	0.076 [0.042]*	0.007 [0.042]
SFSP	-0.027 [0.166]	-0.237 [0.298]	0.139 [0.193]	0.058 [0.032]*	0.030 [0.061]	0.076 [0.034]**
Observations	697	265	432	697	265	432

Notes: OLS coefficients are reported, with heteroskedasticity-robust standard errors in brackets. The row labeled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Sample is all students participating in the STAR program who were registered as of October, 2005, restricted as stated in each panel. All regressions control for high school grade quartile, mother tongue, high school grade and number of courses. Cols (1) and (4) also control for sex.

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

Table 9a: Treatment Effects on First-year Grade, 2SLS Estimates

	Unrestricted			SSP/Any SFP		
	All	Males	Females	All	Males	Females
Program	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: All Students			
Control mean	60.1 (12.9)	60.8 (13.2)	59.7 (12.7)	60.1 (12.9)	60.8 (13.2)	59.7 (12.7)
SSP Participant	-1.60 [1.760]	-2.74 [3.509]	-0.434 [1.868]	-1.62 [1.760]	-2.77 [3.507]	-0.453 [1.866]
Any SFP Participant				0.721 [0.857]	-0.557 [1.466]	1.52 [1.047]
SFP Participant	0.290 [0.925]	-0.952 [1.628]	1.11 [1.111]			
SFSP Participant	1.78 [1.551]	0.304 [2.506]	2.60 [1.970]			
Observations	1561	661	900	1561	661	900
			Panel B: Students with ≥4 Courses			
Control mean	60.8 (12.5)	61.4 (12.9)	60.3 (12.2)	60.8 (12.5)	61.4 (12.9)	60.3 (12.2)
SSP Participant	-0.404 [1.710]	-0.708 [3.301]	-0.206 [1.908]	-0.470 [1.709]	-0.822 [3.296]	-0.256 [1.906]
Any SFP Participant				0.481 [0.865]	-1.11 [1.472]	1.49 [1.065]
SFP Participant	-0.435 [0.955]	-2.45 [1.657]	0.767 [1.166]			
SFSP Participant	2.91 [1.481]**	2.12 [2.441]	3.53 [1.877]*			
Observations	1345	554	791	1345	554	791
			Panel C: Students with at least 5 Courses			
Control mean	62.1 (11.2)	63.4 (10.8)	61.2 (11.3)	62.1 (11.2)	63.4 (10.8)	61.2 (11.3)
SSP Participant	0.578 [2.025]	-3.26 [3.805]	2.83 [2.212]	0.509 [2.023]	-3.26 [3.797]	2.68 [2.206]
Any SFP Participant				0.859 [1.092]	-1.36 [1.898]	1.95 [1.365]
SFP Participant	-0.002 [1.140]	-1.43 [1.738]	0.688 [1.465]			
SFSP Participant	3.06 [2.154]	-1.24 [3.890]	6.05 [2.502]**			
Observations	694	264	430	694	264	430

Notes: Program consent instrumented with invitation to SSP, invitation to SFP, invitation to SFSP. Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Sample is all enrolled students participating in the STAR program with at least one grade as of May, 2006, restricted as stated in each panel. All regressions control for high school grade quartile, mother tongue, and number of courses enrolled. Columns (1) and (4) also control for sex.

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

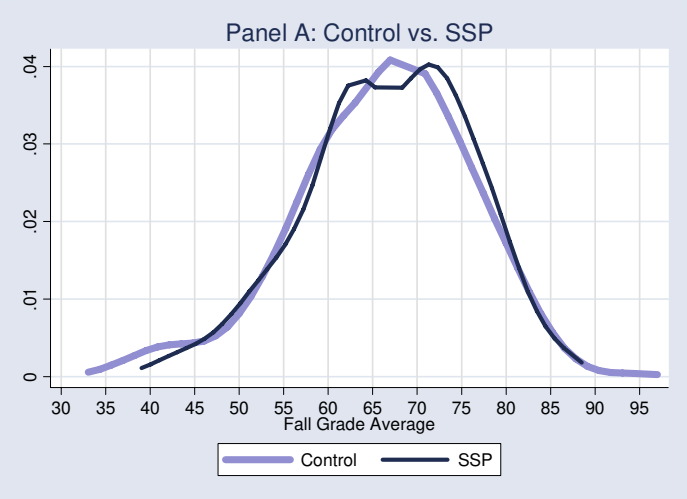
Table 9b: Treatment Effects on Retention, 2SLS Estimates

	Unrestricted			SSP/Any SFP		
	All	Males	Females	All	Males	Females
Program	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: All Students			
Control mean	0.858 (0.349)	0.855 (0.352)	0.860 (0.348)	0.858 (0.349)	0.855 (0.352)	0.860 (0.348)
SSP Participant	-0.044 [0.051]	0.038 [0.081]	-0.103 [0.064]	-0.045 [0.051]	0.038 [0.081]	-0.104 [0.064]
Any SFP Participant				0.039 [0.023]*	0.075 [0.037]**	0.016 [0.030]
SFP Participant	0.035 [0.027]	0.081 [0.040]**	0.005 [0.035]			
SFSP Participant	0.051 [0.037]	0.060 [0.061]	0.046 [0.044]			
Observations	1607	683	924	1607	683	924
			Panel B: Students with ≥4 Courses			
Control mean	0.883 (0.322)	0.874 (0.332)	0.889 (0.314)	0.883 (0.322)	0.874 (0.332)	0.889 (0.314)
SSP Participant	0.003 [0.047]	0.092 [0.074]	-0.056 [0.060]	0.002 [0.047]	0.092 [0.073]	-0.057 [0.059]
Any SFP Participant				0.030 [0.023]	0.064 [0.037]*	0.011 [0.030]
SFP Participant	0.015 [0.027]	0.066 [0.041]	-0.013 [0.036]			
SFSP Participant	0.071 [0.033]**	0.061 [0.060]	0.080 [0.037]**			
Observations	1352	557	795	1352	557	795
			Panel C: Students with ≥5 Courses			
Control mean	0.890 (0.313)	0.883 (0.322)	0.894 (0.308)	0.890 (0.313)	0.883 (0.322)	0.894 (0.308)
SSP Participant	0.053 [0.051]	0.071 [0.082]	0.039 [0.064]	0.052 [0.051]	0.071 [0.082]	0.036 [0.064]
Any SFP Participant				0.045 [0.031]	0.071 [0.049]	0.030 [0.039]
SFP Participant	0.031 [0.035]	0.088 [0.050]*	0.008 [0.046]			
SFSP Participant	0.079 [0.044]*	0.041 [0.082]	0.103 [0.048]**			
Observations	697	265	432	697	265	432

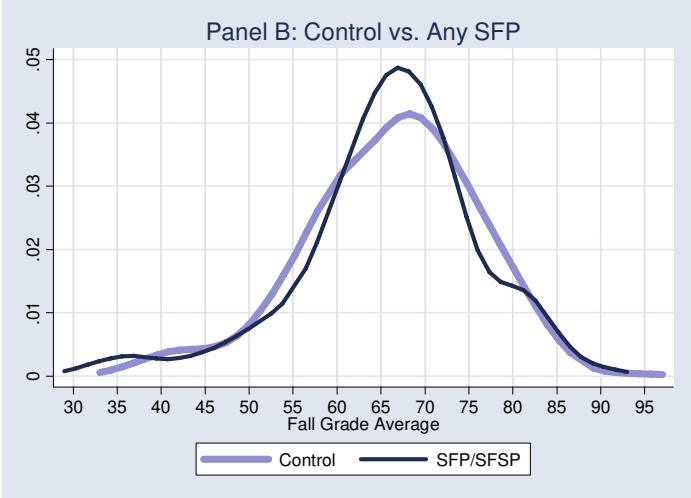
Notes: Outcome is registration for the second year. Program consent instrumented with invitation to SSP, invitation to SFP, invitation to SFSP. Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Sample is all enrolled students participating in the STAR programas of May, 2006, restricted as stated in each panel. All regressions control for high school grade quartile, mother tongue, and number of courses enrolled. Column (1) and (4) also control for sex.

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

Figure 1a: Males' Fall Grade

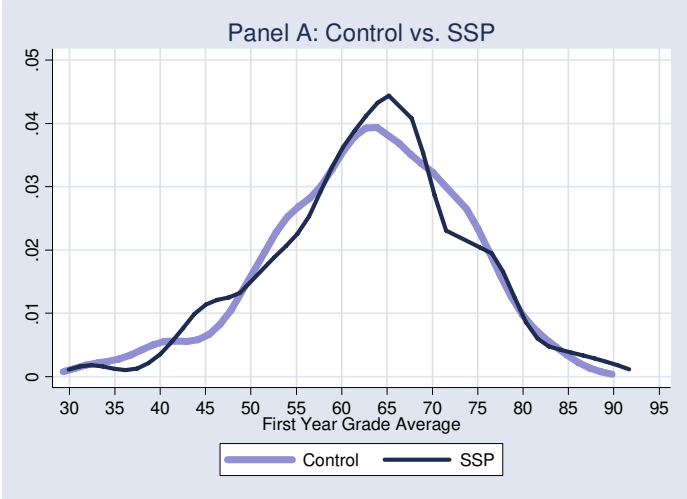


K-S p-value: 0.960

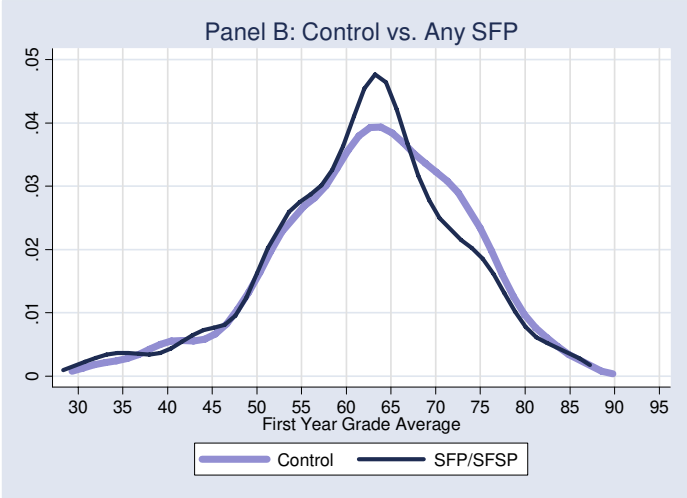


K-S p-value: 0.462

Figure 1b: Males' First-year Grade, Fall Grades Sample

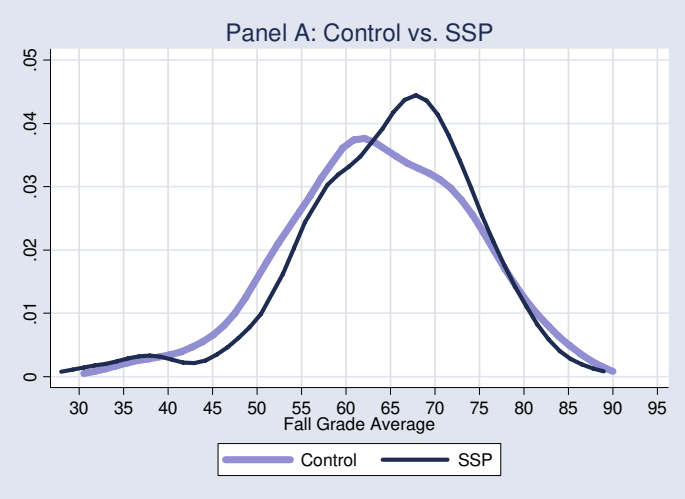


K-S p-value: 0.934

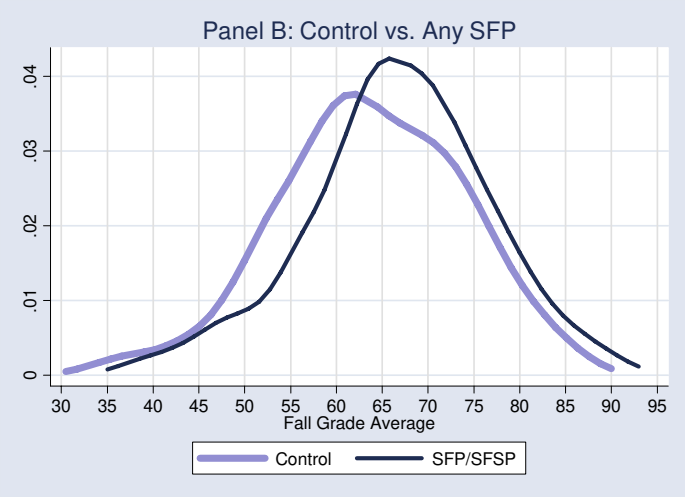


K-S p-value: 0.506

Figure 1c: Females' Fall Grade

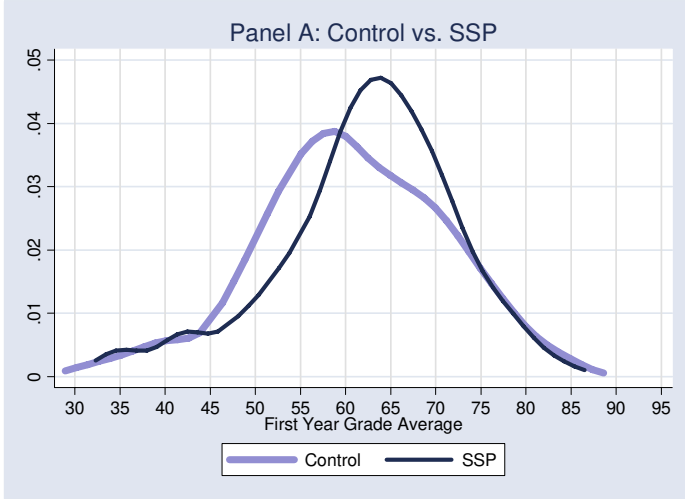


K-S p-value: 0.097

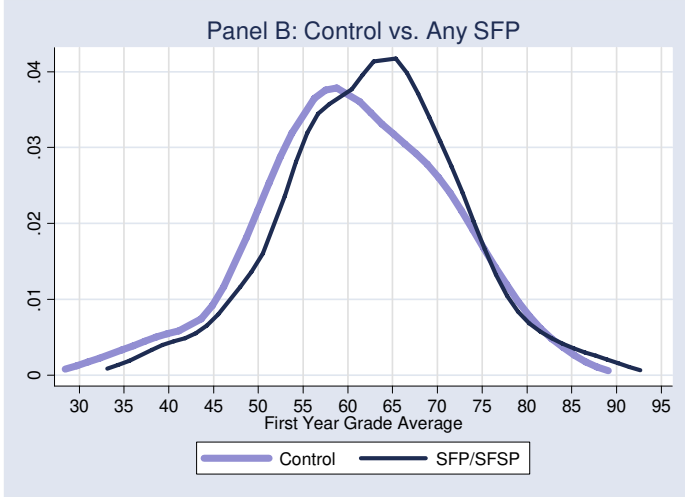


K-S p-value: 0.000

Figure 1d: Females' First-year Grade, Fall Grades Sample



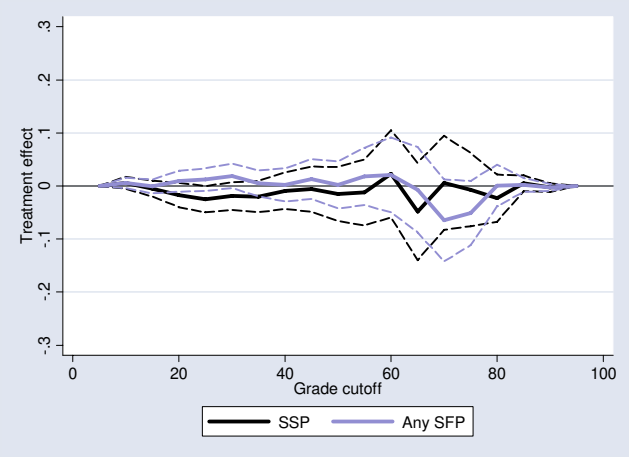
K-S p-value: 0.022



K-S p-value: 0.003

Figure 2a: Treatment Effects by Fall Grades Cutoff, Males

Panel A: SSP vs. control and Any SFP vs. control



Panel B: SFP vs. control and SFSP vs. control

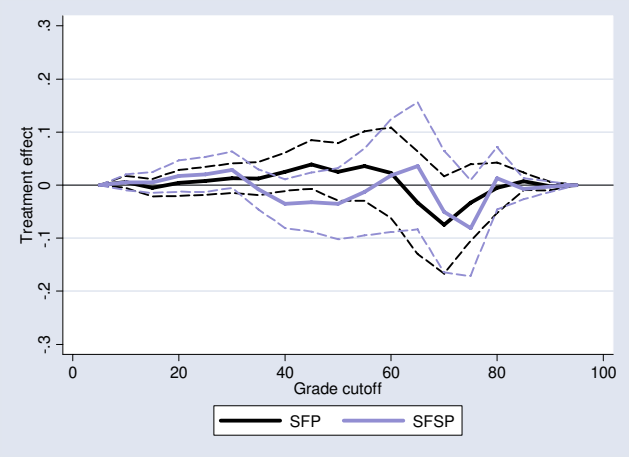
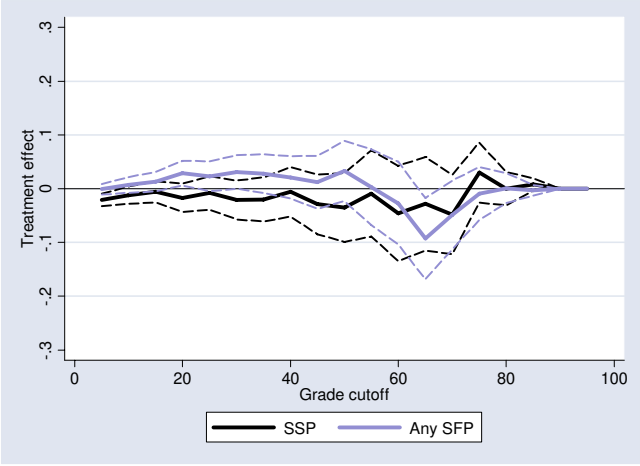


Figure 2b: Treatment Effects by Grade Cutoff, Males with Fall Grades

Panel A: SSP vs. control and Any SFP vs. control



Panel B: SFP vs. control and SFSP vs. control

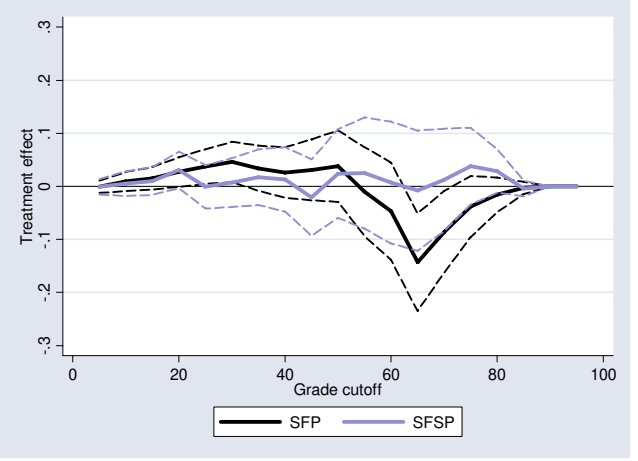
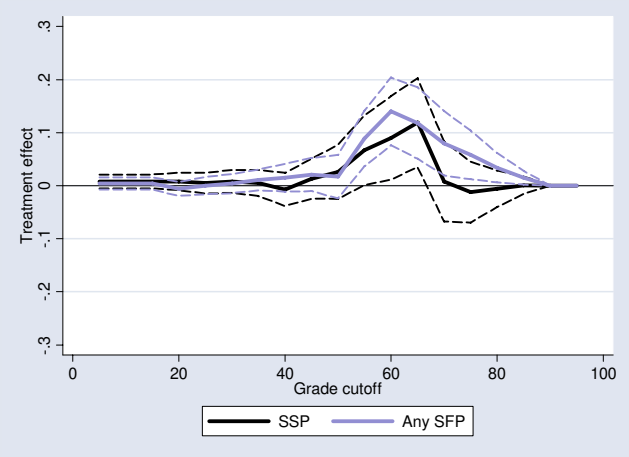


Figure 2c: Treatment Effects by Fall Grades Cutoff, Females

Panel A: SSP vs. control and Any SFP vs. control



Panel B: SFP vs. control and SFSP vs. control

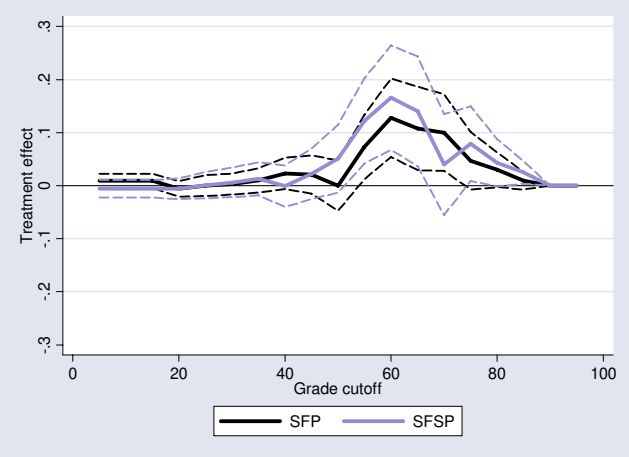
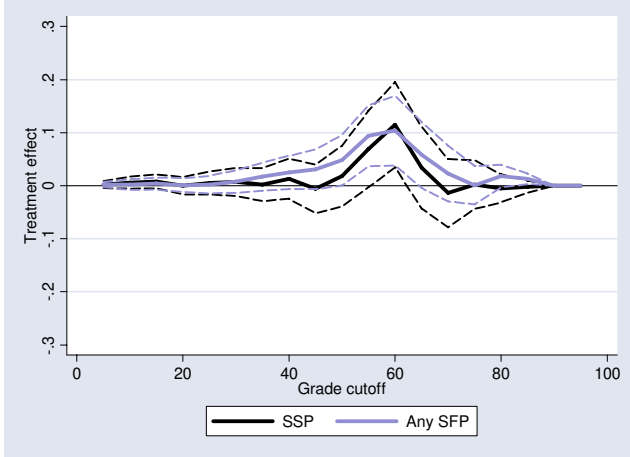


Figure 2d: Treatment Effects by Grade Cutoff, Females with Fall Grades

Panel A: SSP vs. control and Any SFP vs. control



Panel B: SFP vs. control and SFSP vs. control

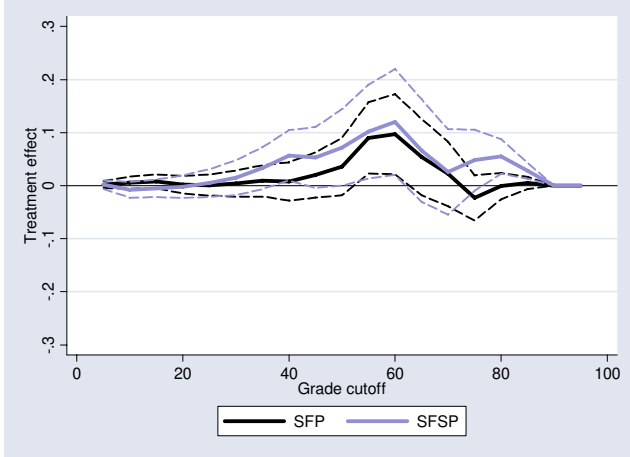
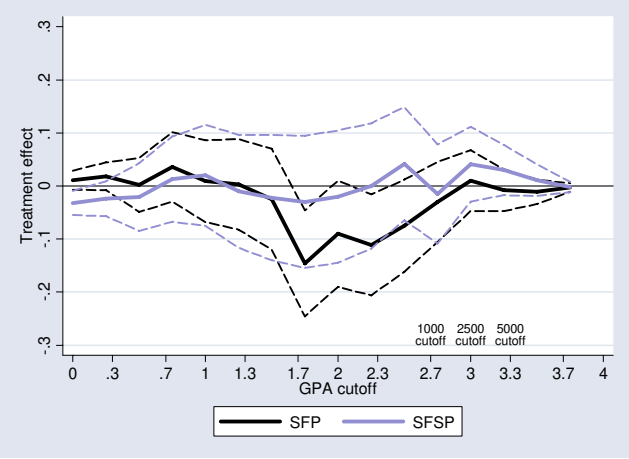


Figure 3a: Treatment Effects by Standardized GPA Cutoff, Males

Panel A: SFP vs. control and SFSP vs. control, Full Sample



Panel B: SFP vs. control and SFSP vs. control, ≥4 courses sample

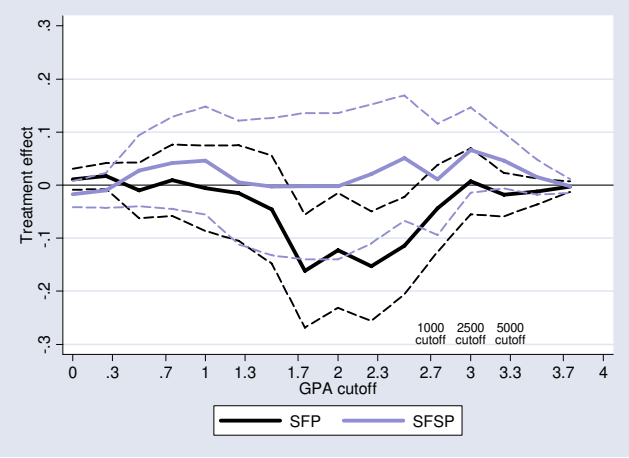
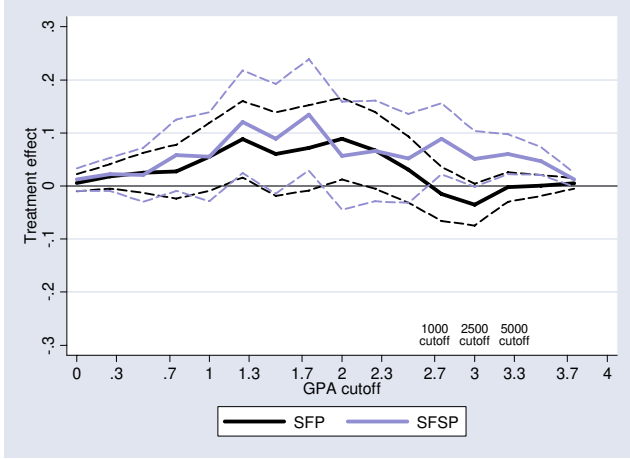
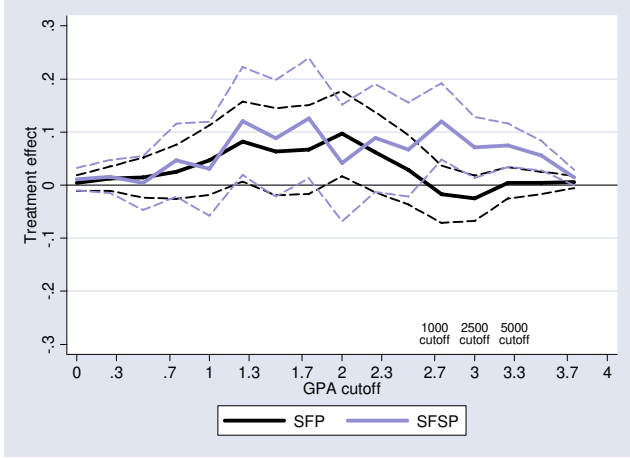


Figure 3b: Treatment Effects by Standardized GPA Cutoff, Females

Panel A: SFP vs. control and SFSP vs. control, Full Sample



Panel B: SFP vs. control and SFSP vs. control, ≥4 courses sample



Appendix: Student Fellowship Program Award Schedule

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