PRELIMINARY AND INCOMPLETE

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

By,

Joshua Angrist
MIT, NBER

Daniel Lang
University of Toronto

Philip Oreopoulos
University of Toronto, NBER

November 2006

This project was sponsored by the Canada Millennium Scholarship Foundation. Special thanks go to Yan Tam-Seguin and Daniela Rusivka, as well as other staff and administrators who assisted with this demonstration project. 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, University of Toronto, Williams College, Harvard University, Statistics Canada, and the Stanford Institute Theoretical Conference (SITE).

Lead them to Water and Pay them to Drink:

A Field Experiment Offering Guidance and Incentives for College Achievement

Abstract

High attrition rates, delayed completion, and poor achievement are growing concerns among colleges and universities in North America. This paper reports on a randomized field experiment involving two strategies designed to improve these outcomes. The results reported here are for approximately 1,600 first-year undergraduate students who participated in the Student Achievement and Retention Project at a large Canadian university. One treatment group was offered peer advising and critical-thinking tutorial services. Another group was offered substantial merit-scholarships for solid, but not necessarily top, first year grades. A third treatment group combined both interventions. While service take-up was low, it was much higher for students offered both programs than for those offered services alone. Females also used services more than males. No program had an effect on grade or retention outcomes for males. However, first-term grades were significantly higher for females in the two scholarship groups. Program effects on females abate somewhat when looking at final, first-year grades and other end-of-year outcomes, but remain significant for females who pre-registered for a full course load. On balance, the results suggest a combination of services and incentives is more promising than either alone, perhaps because financial incentives increase use of services even when the two strategies are not formally linked, or because the combined treatment included more reminders about the scholarship program.

Joshua Angrist
MIT Department of Economcs
angrist@mit.edu

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

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

I. Introduction

Recent years have seen growing interest in policy strategies designed to increase college attendance, 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 that 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 may be an individually rational response to new information about costs and benefits. Nevertheless, policy-makers typically see dropout behaviour and low achievement as undesirable and college administrators invest considerable

[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).

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 (Irvine, 1996). 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, 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 Hope. 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 these new scholarship programs boost college

[2]The National merit program awards 8200 scholarships out of 1.4 million PSAT-takers.

attendance and completion (Dynarski, 2000, 2002), but the evidence is mixed and the cost of these programs is 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 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. 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 upperclass 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, as far as we know, has not been looked at previously using any sort of research design.

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

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

discussion with focus groups in an attempt to understand the findings. Finally, Section VI offers a summary and some conclusions.

Overall, the effects of the STAR intervention were small and not statistically significant from zero. A closer look, however, reveals important effects of the fellowship on females, especially in some subgroups. Females in the fellowship group had markedly better Fall-term grades and somewhat better grades at years end, though the initial boost faded considerably. The effects on females come from the subgroup carrying a full course load. Another important result is that the year-end effects were much 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. Both our quantitative and qualitative findings suggest that a combination of services and incentives is more promising than either alone, perhaps because incentives increase use of services even when the two are not formally linked, or because the combined treatment included more reminders about the scholarship program.

II. Background and Context

The benchmark economic model of schooling-as-human-capital treats educational attainment and student effort as outcomes 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). New information after entering college may make some students update their cost-benefit problem and leave. Some students may dislike studying or benefit from it more than others. Moreover, a number of empirical studies suggest the returns to a partially completed degree are not substantially below the returns to completion in per-year terms (Kane and Rouse, 1995). Viewed in this way, it is not clear why the decision to leave school or perform poorly 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 dropout. 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 (Manski, 1993). 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. Students clearly run into trouble in college 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 college performance and attrition [Lotkowski, Robbins, and Noeth, 2004]. Aware of this fact, many institutions offer an array 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. Girabaldi, *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 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. DeJardins 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. The average number of 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 stipends for high school students 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, ours 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 are 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 upperclass 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.

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 (Halbley, 2004). Several universities assign first year students to an upper-class peer or faculty advisors who provides academic support. Wyckoff (1998) suggests these informal and formal interactions increase persistence. Few colleges, however, have structured as extensive a mentoring program

[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.

[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.

as was offered through the SSP. Peer advisors in our 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.

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 sit in on the course 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]

[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.

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 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.

Students on the satellite campus 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 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.

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

[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.

[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.

and locations were announced often.[9] After the first semester, we also offered bookstore gift certificates to those who attended FSGs and 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 chose not to be involved; these students were not eligible for fellowships even if they met SFP grade targets and therefore should be unaffected by the 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

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

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 D_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 becomes the adjustment originally proposed by Bloom (1984)

$$E[Y_{1i}!Y_{0i}|\ D_i=1] = \{E[Y_i|\ Z_i=1]!E[Y_i|\ Z_i=0]\}\ H\ P[D_i=1|\ Z_i=1].$$

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; Imbens and Angrist, 1994). In the SSP and SFSP, a further distinction might be made between compliance-as-consent and compliance-as-service-use. On the other hand, the availability of services and interest shown by peer advisors (who emailed participants biweekly) is a service received by all SSP and SFSP consenters. Therefore, we make no adjustment for the difference between consent and usage.

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 our background survey.[10] 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

[10] 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.

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. Seventy-seven percent of students in our sample live at home with their parent(s). Only a few of identified this campus as their first choice for post-secondary education studies. Most of the study population worked in high school and the majority plan to work at least part-time while in college. Forty-two percent said they planned to work more than 10 hours a week. Many of the students are immigrants or children of immigrants, as suggested in the fact that 30 percent have a non-English mother tongue.[11] The students' parents, however, are reasonably well-educated; many have college degrees (though it should be noted that these are student-reported parental schooling measures). All claim to put a priority on high grades. Interestingly, 64 percent said they wanted more education than a bachelor's degree and 83 percent said they intend to complete their undergraduate program in 4 years. Among those who entered in 2001, only 38 percent completed a degree this quickly. Overall, the first year retention rate was about 13 percent, and the six-year graduation rate about 70 percent.

Merit scholarship programs like ours may affect course enrolment decisions and/or the selection of courses by treated students. We tried to minimize this behaviour by contacting treated students for the first time only after initial course selection. It is difficult to change courses later because many fill to capacity and because of 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

[11] Few students are French-speaking. Most of the non-English speakers are from Asia-Pacific or South-Asia.

difficult). The estimates indicate that the STAR demonstration did not affect initial course load or selection.

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.[12] Consent serves as an indicator of student awareness and interest.[13] 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 reports treatment-control differences and estimates from regression models including two 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 were much more likely to consent to participate in STAR than were males in each of the three treatment groups. 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 and 81 percent of males having consented. Similarly, when offered both services and fellowships in the SFSP, 84 percent of females consented while 71 percent of males consented.

---

[12] General information to the control group was provided if they inquired about the program. Few did.
[13] Insert more about logisitcs in initial mailings and follow-ups for report here.
[14] In these regressions and elsewhere in paper, columns labeled "Basic controls," report estimates of the coefficient on assignment-group dummies in models that control for sex, mother tongue, high school grade quartile, and number of courses at pre-registration. These are variables from administrative data. Columns labeled "All controls," add the responses to five survey questions, whether UTM was the subject's first-choice, hours/week of study in high school, hours/week of planned market work at UTM, parents' education.

The pattern of service use shows differences by treatment arm and sex similar to that observed for 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), both in the Fall and in the Spring. Females were also much more likely to use services than males. For example, 16 percent of all students offered services only used some kind of service (peer mentoring or supplemental instruction) in both Fall and Spring, but service use was close to 30 percent in the SFSP. Fall service use by females was 35 percent, while Fall service use for males was about 20 percent. These estimates appear in columns 3-6. 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 the opportunity to win a fellowship motivated students to use services throughout the year. This may signal increased effort throughout the year as well.

Overall service-use rates were somewhat higher than the termwise rates reported in Table 3a. This can be seen in columns 1 and 2 of Table 3b, which reports the probability SSP and SFSP students used services at any time. The highest use rate - 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 the supplemental instruction treatment 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). Use rates for both types of service 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 to offer FSGs in every course in which STAR participants were enrolled (86 percent of subjects attended at least one course incorporating an FSG), and by the fact the we offered services to individual students as opposed to entire classrooms. On the other hand, supplemental instruction take-up rates in the STAR

demonstration are not out of line with take-up rates for supplemental instruction reported elsewhere in the literature (REFERENCE).

There are important differences in both consent rates and service use by students' academic background and pre-registration course load. These differences are explored in Table 3c, which reports estimates of main effects and interaction terms in the model

$$y_i = X_i'B + 6_iSSP_i + 8_iSFP_i + :_iSFSP_i + ,_i, \quad (1)$$

where

$$6_i = 6_{0i} + 6_{1i}LOWHS_i + 6_{2i}LOAD4_i$$

$$8_i = 8_{0i} + 8_{1i}LOWHS_i + 8_{2i}LOAD4_i$$

$$:_i = :_{0i} + :_{1i}LOWHS_i + :_{2i}LOAD4_i$$

where y_i is 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., $6_{0i} + 6_{1i}$; $6_{0i} + 6_{1i} + 6_{2i}$).

One of the largest take-up differential documented in Table 3c is a lower consent rate and reduced likelihood of service use by males with bottom-quartile 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 the interactions for females are not significantly different from zero. Another differential documented in Table 3c is the consistently higher take-up rate and rate of SSP service use 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. Effects on Course Credit and Year-two Enrollment

Two primary goals of the STAR demonstration were to increase the number of credit-units (courses) students complete successfully in freshman year, and to increase the likelihood that they go on to second year without dropping out of school. The treatments piloted in STAR had no discernible effect on either of these outcomes.

The effects of the SSP, SFP, and SFSP on first-year credits and second-year registration are presented in Table 4. The table reports regression estimates from a version of equation (1) without interaction terms. The average number of credits completed in the control group is 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 course completion or retention as a function of the three STAR treatments, as can be seen in Panel A of Table 4. The estimates for females, in Panel C, show mostly positive effects on credit-units ranging from .12-.15 in column 3, but none of these are much larger than the corresponding standard errors. The estimated effect on credit units for males are mostly negative, thought two specifications generate small but marginally significant positive effects of the SFP on retention.

The fact that there is little evidence of an effect on credits or retention raises the question of whether services or fellowships had any effect we can measure. One possible explanation for the lack of an effect on these key outcomes is that students were unaware of the intervention, or failed to react to it in any way. On the other hand, the program may have generated a reaction that did not translate into detectable effects on retention or credits. As it turns out, the two fellowship treatments increased achievement for females, at least in the first term of freshman year. There is also some evidence of an achievement boost for a subgroup of females who pre-registered for at least 4 courses, the minimum load required to be eligible for a fellowship. These results are discussed in the next section.

V. Achievement Effects

We begin the analysis of achievement by looking at students' average grades at the end of the Fall semester and at the end of the first year. The grade outcome is a credit-weighted average on a 0-100 grading scale. The Fall grades outcome provides 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 be carrying 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 5, which report treatment effects on the full sample estimated with varying sets of controls. Most of the SFP and SFSP effects 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.

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 5. 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 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), while the corresponding estimate of the SFSP effect on females is 3.3 (s.e.=1.5). These estimates from

column 5 of Panel C in Table 5. 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 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 and first-year grades for females in the Fall grades sample. There is also at least a marginally significant distribution shift associated with the SSP, though as Table 5 shows, this did not translate into a significant mean effect (except in one specification with the full set of survey controls).

The pointwise confidence intervals for distribution treatment effects plotted in Figure 2 give a picture of 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 $\{1[Y_i>c]; \quad c=.05, .1, .15 \ldots .95\}$ 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 5 and is 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 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 markedly weaker than in the Fall grades sample. This is apparent in Panel A of Table 6, which reports estimates 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 5, 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, are also insignificant; the effect on females in this specification is 1.4 (s.e.=.93). The difference between the full and Fall-grades sample is highlighted by Panel B, which shows results from the same specifications using the Fall grades sample. The any-SFP effect on females in the Fall-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 Fall grades sample consists almost entirely of students who pre-registered for four or five courses per term. 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 coursess have therefore signalled a stronger commitment to their studies than those taking a lighter load.

Estimates in the sub-samples of students taking or four or five courses, reported in Panels C and D of Table 6, support the notion that part of the explanation for the difference in results between the Fall and full samples is the fact that those with Fall grades took more courses. About half of the females pre-registered for a full load of course (430 out of 900), XX of whom have a Fall grade. 88 percent (791) of females pre-registered for 4 or more courses, XX of these have Fall grades. 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. On the other hand, the Any-SFP treatment is estimated to have had a 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, the full-year 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 result, but this may be fortuitous. On the other hand, other sample-section rules, motivated by the fact that only students with a higher course load could qualify for fellowships, and may otherwise have been more likely to benefit from and motivated to win fellowships, generate significant results for females offered a combination of fellowships and services.

The remainder of this section looks briefly at effects on outcomes other than grades, 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. Finally, we report results from a two-stage least squares procedure that adjusts the 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 7.[15]

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

[15] Table 7 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 result in a somewhat more generous award rate.

Table 7, 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 mean effects on 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 Table 7, 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 7, 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. Also marked on the figures are the cut-offs for fellowship eligibility (the figure uses GPA's adjusted so that students from all high school grade quartiles have approximately the same cutoff). 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 effects on males in the middle of the distribution, while the estimates for females show mostly positive effects, some significant. Some of the significant SFSP effects are the neighbourhood of the award cut-offs but shifts 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 models with interaction terms are essentially the same as equation (1), except that the second interaction is a dummy for planning to work more than 5 hours per week (the approximate median of this variable). The results are reported in Table 8, 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 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. A work interaction might arise if students who plan to hold down a part-time job find the fellowships especially useful.

In practice, there is no clear pattern of treatment-effect variation with either the GPA or work covariates. There are no significant interactions in the SSP treatments for males or females. Although some of the interactions of the SFP effect with low GPA 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, for females in the four-course sample. The total effects in this specification, reported in column 6, are also marginally significant.

C. Two-stage least squares estimates

second stage:

$$y_i = X_i'* + \forall SSP\_CONSENT_i + \exists SFP\_CONSENT_i + (SFSP\_CONSENT_i + 0_i,$$

Table 9

VI. Student Reports (Focus group post-mortem)

VII. Summary and Conclusions

References

Akerlof, George and Rachael Kranton, "Identity and Schooling: Some Lessons for the Economics of Education," Journal of Economic Literature, No. 40, Vol. 4 (December), 2002, pp. 1167-1201.

Angrist, Joshua and Lavy, Victor, "The Effect of High School Matriculation Awards: Evidence from Randomized Trials." NBER Working Paper 9389, December 2002, pp 1-38

Arendale, David R. "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," University of Missouri-Kansas City mimeo, 2006.

Ariely, D. and K.Wertenbroch (2002): ''Procrastination, Deadlines, and Performance: Self-Control by Precommitment,'' Psychological Science, 13, 219-224.
Arnold, Jane (1996) "Student Retention: Why do we keep losing them?", The NEA Higher Education Journal, Vol. ?? pp. 131-138

Ashworth, K., et al., (2001), Education Maintenance Allowance: The First year, A Quantitative Evaluation, Department for Education and Evaluation Research Brief257, May.

Bailey, Thomas R., and Mariana Alfonso, "Paths to Persistence: An Analysis of Research on Program Effectiveness at Community Colleges," Lumina Foundation For Education New Agenda Series, Vol. 6, No. 1., January 2005.

Bishop, John (2002) "A Prospective Policy Evaluation of the Michigan Merit Award Program," mimeo preparted for 'Taking Account of Accountability: Assessing Politics and Policy," Kennedy School of Government, Harvard University.

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

Brock, Thomas, and Lashawn Richburg-Hayes, "Paying for Persistence: Early Results of a Louisiana Scholarship Program for Low-Income Parents Attending Community College," MDRC Report, May, 2006.

Cameron, J., K. M. Banko, et al. (2001). "Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues." The Behavior Analyst 24: 1-44.

Congos, D. H., & Schoeps, N. (2003). Inside Supplemental Instruction (SI): One model of what happens that improves grades and retention revisited. *Journal of Student Centered Learning, I*(13), 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, "The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's hOPE Program," Journal of Labor Economics, Vol. 24, No. 4, October 2006, pp. 761-786.

Dale, P. M. and Zych, T. (1996) "A successful college retention program," College Student Journal, Vol. 30, Nol. 3, pp. 354-360.

Deci, E. L. (1971). "Effects of Externally Mediated Rewards on Intrinsic Motivation." Journal of Personality and Social Psychology 18: 105-115.

Deci, E. L., R. Koestner, et al. (1999). "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." Psychological Bulletin 125(627-668).

Dynarski, S. (2003). "The Consequences of Merit Aid." NBER Working Paper #9400.

Dynarski, S. (2005). "Building the Stock of College-Educated Labor," Mimeo.

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

Fudenberg, Drew, and David K. Levine, "A Dual Self Model of Impulse Control," Mimeo, 2005.

Grant-Vallone, Elisa, Kelly Reid, Christine Umali, and Edward Pohlert, "An Analysis of the Effects of Self Esteem, Social Support, and Participation in Student Support Services on Students' Adjustment and Commitment to Collge, " Journal of College Student Retention: Research, Theory and Practice, Vol. 5, No. 3, 2003-2004, pp. 255-274.

Grossman, J. B., & Tierney, J. P. (1998). Does mentoring work? An impact study of the Big Brothers Big Sisters program. *Evaluation Review, 22*(3), 402–425.

Heil, S.H., Tidey, J.W., Holmes, H.W., and Higgins, S.T. (2004). Effects of contingent payments on abstinence and withdrawal among cigarette smokers in their usual environments. Nicotine and Tobacco Research, 6, 471-479

Hensen, Kari A. and Mack Shelley, The Impact of Supplemental Instruction: Results from a large, public, Midwestern university,the Journal of College Student Development, 44(2), 250-259 (2003).

Higgins, S.T., Wong, C.J., Badger, G.J., Ogden, D.E., & Dantona, R.L. (2000). Contingent reinforcement increases cocaine abstinence during outpatient treatment and one year of follow-up. Journal of Consulting and Clinical Psychology, 68(1), 64-72.

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

Kremer Michael, Miguel, Edward and Thornton, Rebecca. , CID Working Paper No. 109, October 2004, pp 1-42

Kruglanski, A., I. Friedman, et al. (1971). "The Effect of Extrinsic Incentives on Some Qualitative Aspects of Task Performance." Journal of Personality and Social Psychology 39: 608-617.

Lavy, Victor, and Analia Schlosser, "Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits" *Journal of Labor Economics,* Vol. 23 No. 4, October 2005

Lepper, M., D. Greene, et al. (1973). "Undermining Children's Interest with Extrinsic Rewards: A Test of the 'Overidentification Hypothesis." Journal of Personality and Social Psychology 28: 129-13

Leuven, E., H. Oosterbeek and B. van der Klaauw, 20035, The effect of financial rewards on students' achievement: Evidence from a randomized experiment, Working paper.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): ''Projection Bias in Predicting

Loewenstein, George, and Ted O'Donoghue, "Animal Spirits: Affective and Deliberative Processes in Economic Behavior," Mimeo, 2005.

Lotkowski, V. A., Robbins, S. B., & Noeth, R. J. (2004). The role of academic and non-academic factors in improving college retention. Washington, DC: ACT policy report. (ERIC Document Reproduction Service No. ED485476)

Loury, Linda Datcher and David Garman. "College Selectivity and Earnings." Journal National Center for Education Statistics (2003) *Remedial Education at Degree-Granting*

O'Donoghue and Rabin (2005), "Incentives and Self-Control", Mimeo, 2005.

O'Donoghue, T. and M. Rabin (1999), ''Doing It Now or Later,'' American Economic Review, 89, 103-124.

Ogden, P., Thompson, D., Russell, H. A., & Simons, C. (2003). The short- and long- term impact of supplemental instruction. Journal of Developmental Education, 26(3), 2-9.

Oreopoulos, Philip (2005) “Do Dropouts Drop Out Too Soon? Wealth, Health, and Happiness from Compulsory Schooling" Mimeo.

Pantages, Timothy, J., and Carol F. Creedon (1978) “Studies of College Attrition: 1950-1975,” Review of Educational Research, Winter 1978, Vol. 48, No. 1, pp. 49-101.

Peltier, Gary L., Rital Laden, and Myrna Matranga (1999), “Student Persistence in College: A Review of Research,” Journal of College Student Retention, 1999/2000, Vol. 1, Nol. 4, pp. 357-375

Postsecondary Institutions in Fall 2000. Washington DC: Department of Education.
Taylor, K., with Moore, W.S., MacGregor, J., and Limblad, J. (2003), “Learning Community Research and Assessment: What we know now” National Learning Communities Project Monograph Series, Olympia, WA: The Evergreen State College, Washington Center for Improving the Quality of Undergraduate Education, in cooperation with the American Association for Higher Education.

Tinto, Vincent. (1993) Leaving college: Rethinking the causes and cures of student attrition (2nd ed.) Chicago: University of Chicago Press

Appendix

COST BENEFIT ANALYSIS

Appendix: Student Fellowship Program Award Schedule

Previous High School Grade Avg. Quartile	$1,000 for reaching a grade pt. average of top 4 credits above:	(for only ½ in SFP) $2,500 for reaching a grade pt. average of top 4 credits above:	$5,000 for reaching a grade pt. average of top 4 credits above
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-)

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]	0.7 (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
UTM first uni. choice	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	0.023	-0.025	0.72		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
Family background variables								
Mother tongue is English	0.701	0.019 [0.032]	0.007 [0.032]	0.046 [0.040]	0.5 (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	
Outcomes								
Fall mark	63.8 {12.1}	0.37 [0.927]	2.01 [0.912]	1.75 [1.178]	2.08 (0.102)	1397	63.8 {12.1}	1397
Spring mark	61.2 {15.0}	-0.861 [1.185]	-0.204 [1.162]	0.091 [1.512]	0.19 (0.905)	1309	61.1 {15.0}	1265
Overall first-year mark	60.1 {12.9}	-0.746 [0.949]	0.322 [0.937]	0.842 [1.189]	0.50 (0.684)	1561	60.2 {12.9}	1397
Official GPA	1.77 {0.922}	-0.015 [0.067]	-0.001 [0.066]	0.062 [0.084]	0.22 (0.880)	1538	1.78 {0.910}	1374
Number of credits completed	3.95 {0.921}	0.061 [0.133]	-0.011 [0.131]	-0.088 [0.167]	0.79 (0.499)	1561	7.96 {1.85}	1397
Number of math and sci. credits completed	1.10 {1.21}	0.080 [0.088]	0.103 [0.087]	-0.180 [0.110]	1.90 (0.128)	1561	1.13 {1.22}	1397

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 (1)	Basic controls (2)	All controls (3)	No controls (4)	Basic controls (5)	All controls (6)	No controls (7)	Basic controls (8)	All controls (9)
Control group mean		0.965 (0.183)			0.942 (0.233)			0.844 (0.363)	
Offered SSP	0.019 [0.010]	0.019 [0.009]	0.008 [0.008]	-0.006 [0.017]	-0.004 [0.011]	0.006 [0.010]	-0.004 [0.026]	-0.002 [0.022]	0.001 [0.022]
Offered SFP	0.019 [0.010]	0.012 [0.008]	0.017 [0.005]	0.026 [0.013]	0.014 [0.011]	0.02 [0.010]	0.032 [0.024]	0.028 [0.022]	0.021 [0.024]
Offered SSP and SFP	-0.005 [0.017]	0.003 [0.014]	-0.001 [0.012]	-0.029 [0.024]	-0.013 [0.017]	-0.007 [0.015]	-0.051 [0.035]	-0.019 [0.031]	-0.009 [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 University of Toronto at Mississauga (UTM) 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 University of Toronto at Mississauga (UTM) 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: Boys			
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: Girls			
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 University of Toronto at Mississauga (UTM) 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: 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.

* 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: Boys			
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: Girls			
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 University of Toronto at Mississauga (UTM) 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 controls for sex.

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

Table 3c: Interactions in Service take-up effects

	STAR Consent			Used SSP Services		
Program	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (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 University of Toronto at Mississauga (UTM) 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 5: 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: Boys			
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: Girls			
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 University of Toronto at Mississauga (UTM) 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: 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. Panel A "Basic" and "All" regressions also control for sex.

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

Table 6: Treatment Effect on First-year Grade (Credit-weighted)

	By type			Pooled		
Program	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (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.)

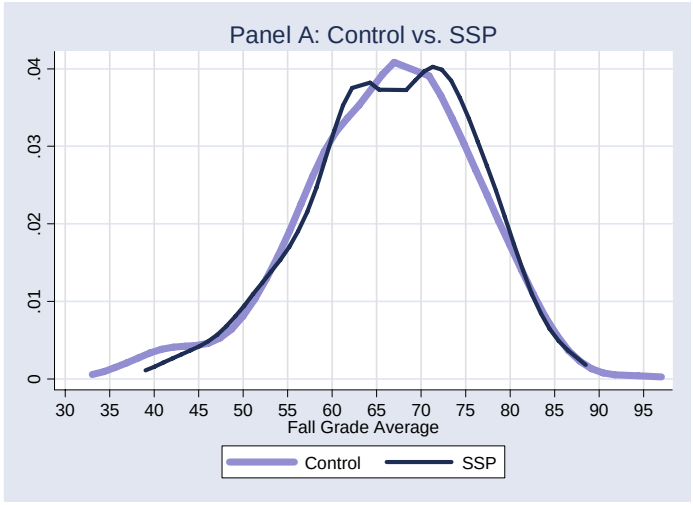
Table 6: Treatment Effect on First-year Grade, cont.

	By type			Pooled		
Program	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (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

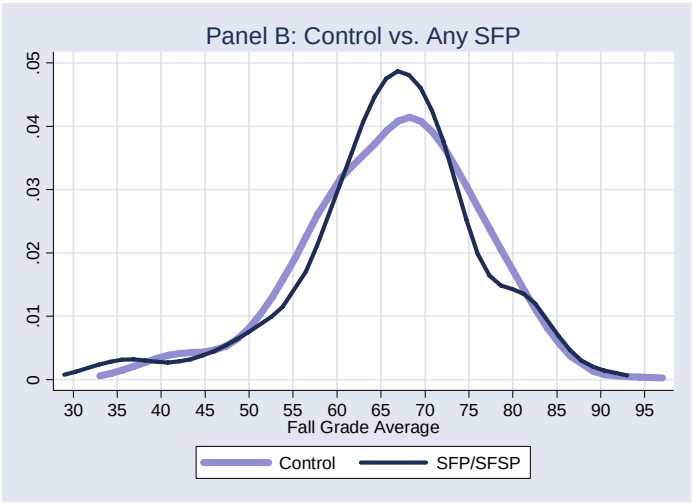
Notes: Heteroskedasticity-robust standard errors in brackets. The row labelled control mean reports the average outcome in the control group, with the corresponding standard deviation in parentheses below. Sample is all enrolled University of Toronto at Mississauga (UTM) students participating in the 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%

Figure 1a: Boys' Fall Grade

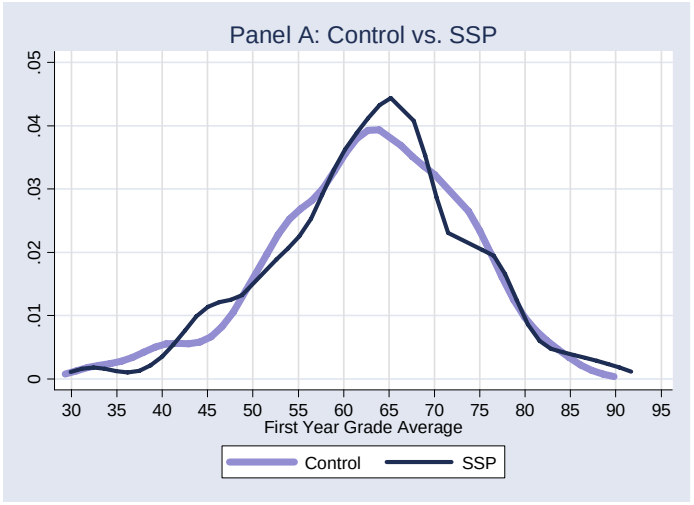


K-S p-value: 0.960

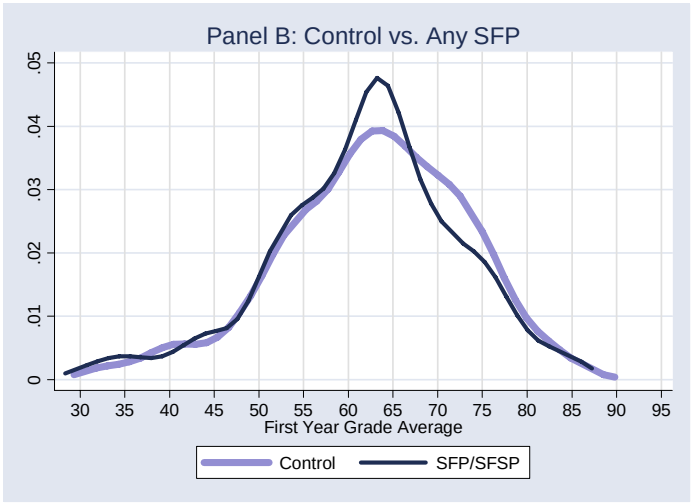


K-S p-value: 0.462

Figure 1b: Boys' Full-year Grade, Fall Grades Sample



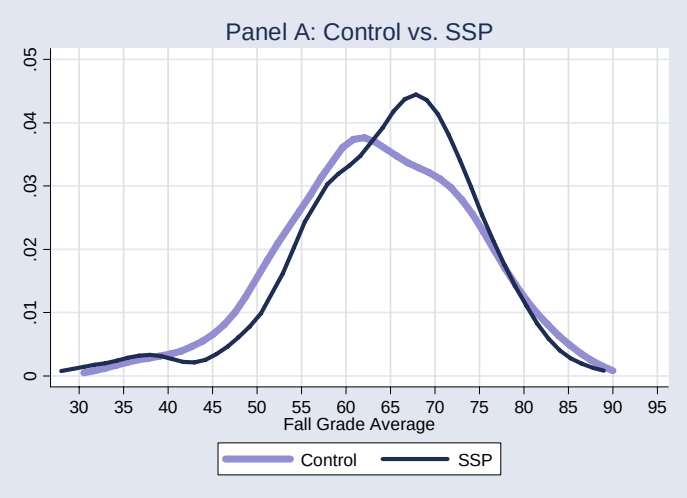
K-S p-value: 0.934



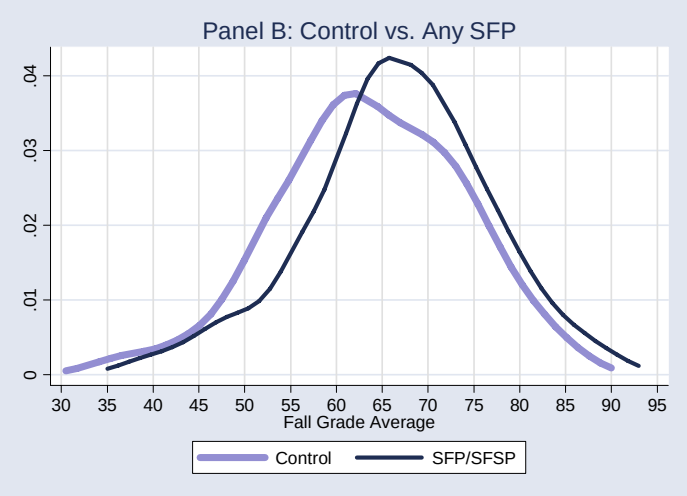
K-S p-value: 0.506

Notes: Plots report smoothed kernel densities of fall term or full-year grades. K-S p-value is the Kolmogorov-Smirnov tests of the equality of the two distributions plotted in each figure.

Figure 1c: Girls' Fall Grade

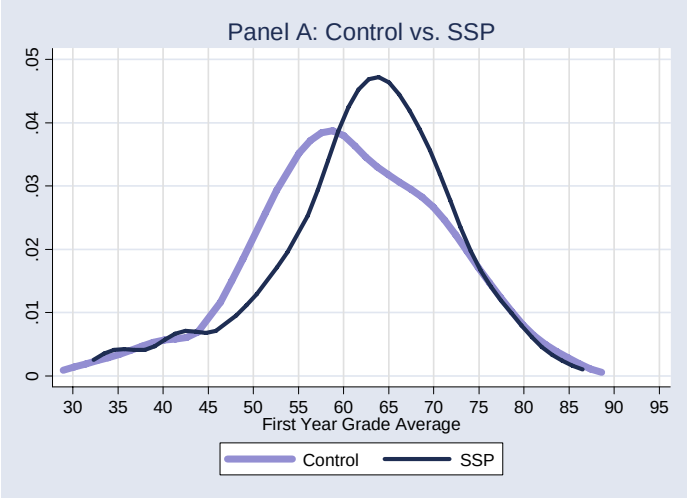


K-S p-value: 0.097

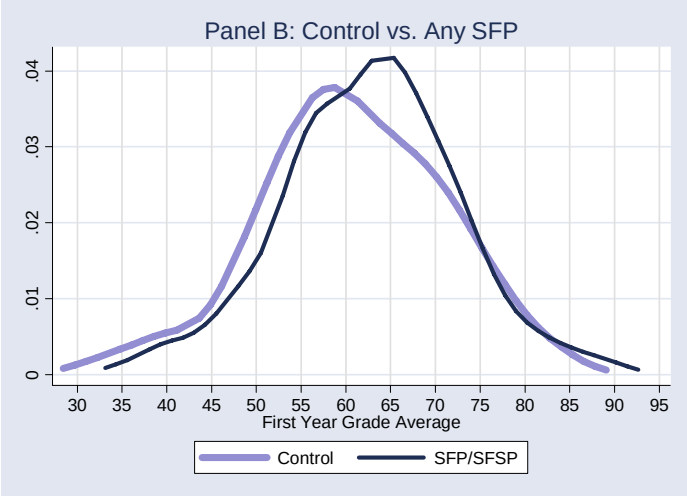


K-S p-value: 0.000

Figure 1d: Girls' Full-year Grade, Fall Grades Sample



K-S p-value: 0.022

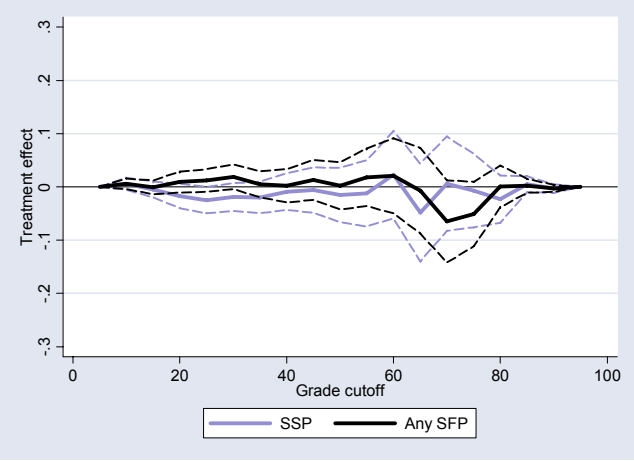


K-S p-value: 0.003

Notes: Plots report smoothed kernel densities of fall term or full-year grades. K-S p-value is the Kolmogorov-Smirnov tests of the equality of the two distributions plotted in each figure.

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

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



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

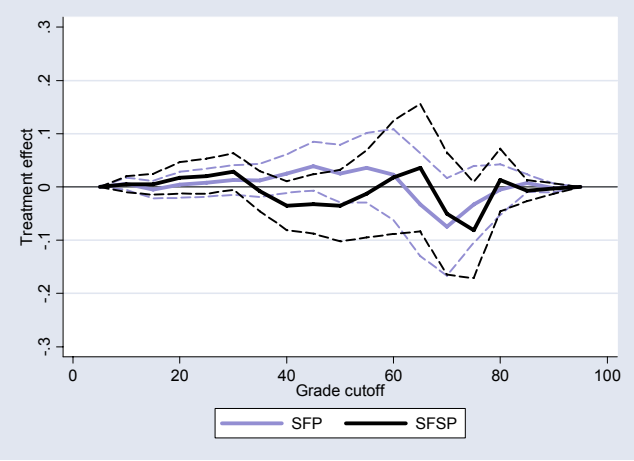
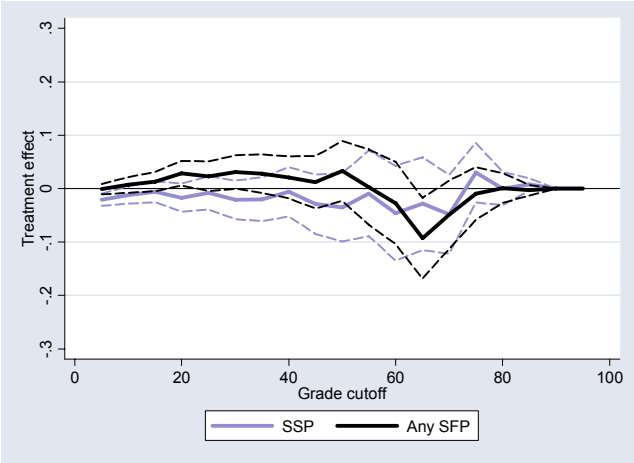
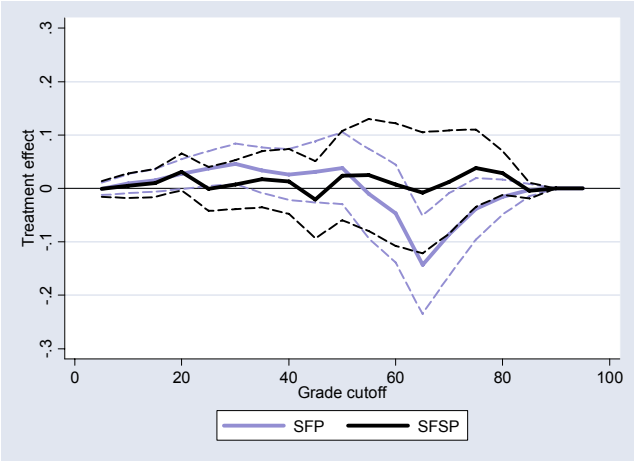


Figure 2b: Treatment Effects by Full-year Grade Cutoff, Boys with Fall Grades

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



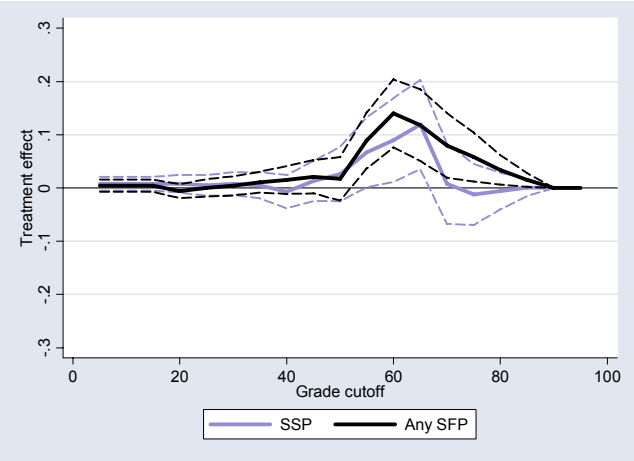
Panel B: SFP vs. control and SFSP vs. control



Notes: Each panel in 2a plots the coefficients on the treatment group indicators from a regression of **1**(fall grade≥x) on treatment and controls for high school grade quartile, mother tongue and number of credits enrolled. Panel 2b plots the results of regressions where **1**(first-year grade≥x) is the LHS variable. Heteroskedasticty-robust 90% confidence intervals are indicated with dashed lines.

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

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



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

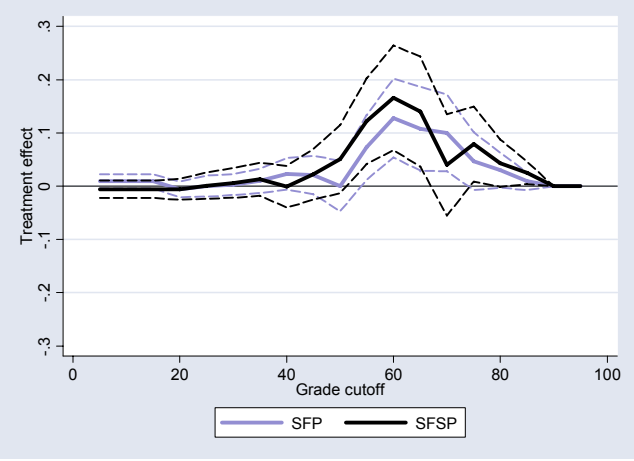
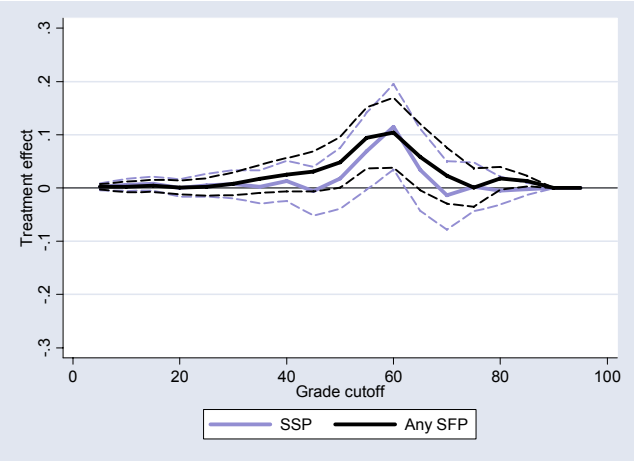
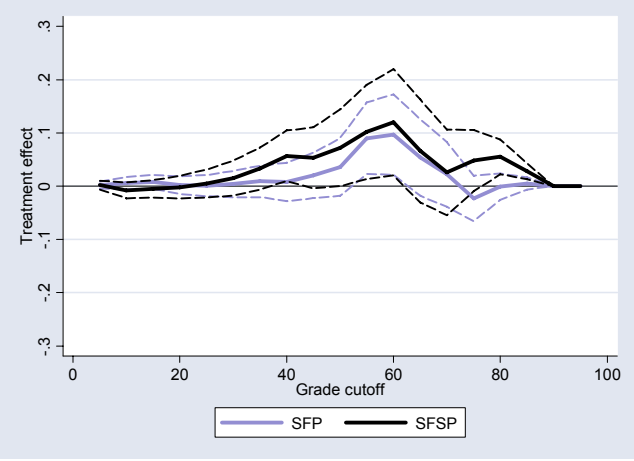


Figure 2d: Treatment Effects by First-year Grade Cutoff, Girls with Fall Grades

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



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



Notes: Each panel in 2c plots the coefficients on the treatment group indicators from a regression of **1**(fall grade≥x) on treatment and controls for high school grade quartile, mother tongue and number of credits enrolled. Panel 2d plots the results of regressions where **1**(first-year grade≥x) is the LHS variable. Heteroskedasticty-robust 90% confidence intervals are indicated with dashed lines.

Table 7: Treatment Effect on Fellowship Eligibility, Fall Grades Sample

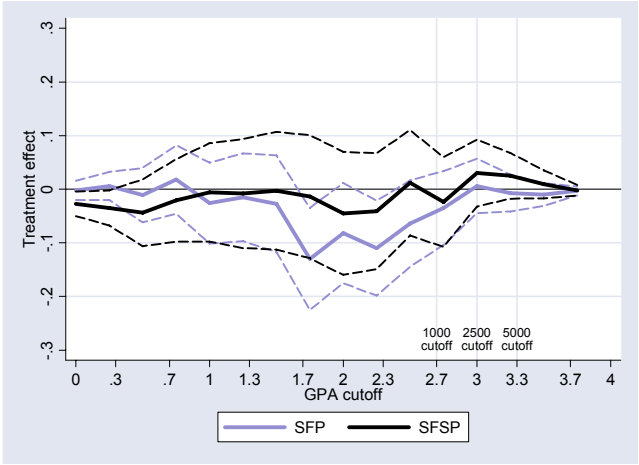
	Pooled								
	$1,000+			$2,500			$5,000		
Program	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (6)	All (7)	Boys (8)	Girls (9)
	Panel A: Students with Grades								
Control mean	0.196 (0.397)	0.230 (0.421)	0.171 (0.377)	0.114 (0.318)	0.141 (0.349)	0.094 (0.292)	0.062 (0.242)	0.074 (0.263)	0.053 (0.225)
SSP	-0.036 [0.026]	-0.071 [0.040]	-0.014 [0.034]	-0.025 [0.021]	-0.065 [0.031]*	0.001 [0.028]	-0.029 [0.014]*	-0.038 [0.023]	-0.024 [0.018]
SFP	0.007 [0.028]	-0.084 [0.039]*	0.072 [0.039]	-0.052 [0.019]**	-0.085 [0.027]**	-0.025 [0.025]	-0.032 [0.014]*	-0.043 [0.020]*	-0.021 [0.018]
SFSP	0.031 [0.036]	-0.001 [0.056]	0.057 [0.048]	0.016 [0.029]	-0.054 [0.038]	0.073 [0.043]	0.023 [0.024]	-0.02 [0.031]	0.058 [0.036]
Observations	1561	661	900	1561	661	900	1561	661	900
	Panel B: Students with ≥4 Courses								
Control mean	0.222 (0.416)	0.266 (0.443)	0.189 (0.392)	0.129 (0.335)	0.164 (0.371)	0.103 (0.304)	0.070 (0.255)	0.085 (0.279)	0.059 (0.236)
SSP	-0.033 [0.029]	-0.078 [0.046]	-0.006 [0.038]	-0.024 [0.023]	-0.074 [0.036]*	0.008 [0.031]	-0.03 [0.016]	-0.042 [0.027]	-0.023 [0.020]
SFP	0.004 [0.031]	-0.1 [0.046]*	0.07 [0.042]	-0.058 [0.021]**	-0.11 [0.030]**	-0.022 [0.027]	-0.032 [0.015]*	-0.046 [0.023]*	-0.02 [0.020]
SFSP	0.029 [0.043]	-0.006 [0.068]	0.056 [0.055]	0.027 [0.036]	-0.059 [0.048]	0.092 [0.050]	0.032 [0.029]	-0.018 [0.038]	0.070 [0.042]
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 University of Toronto at Mississauga (UTM) students participating in the STAR program with at least one grade as of May, 2006. Sample in Panel B is all UTM 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%

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

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



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

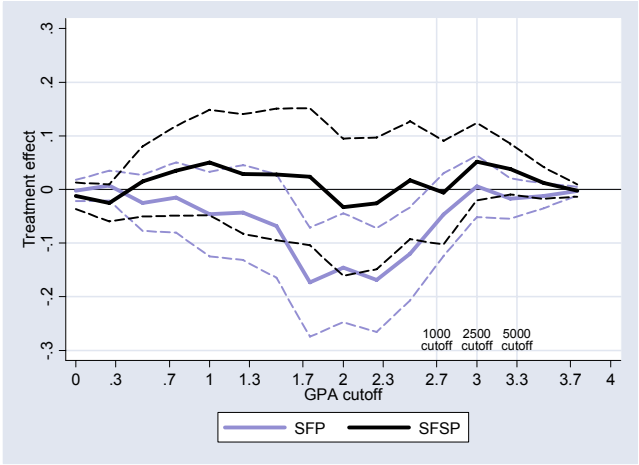
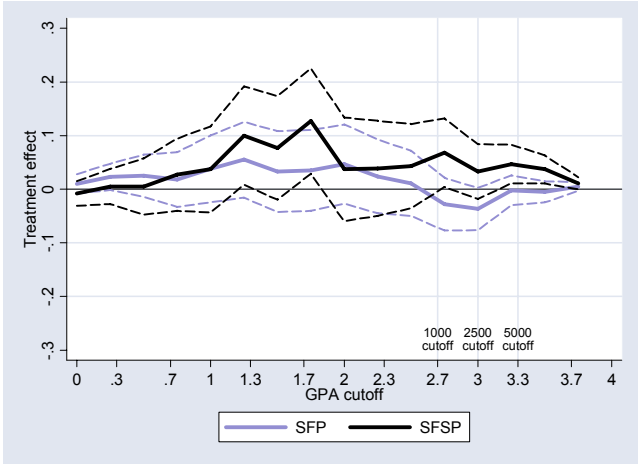
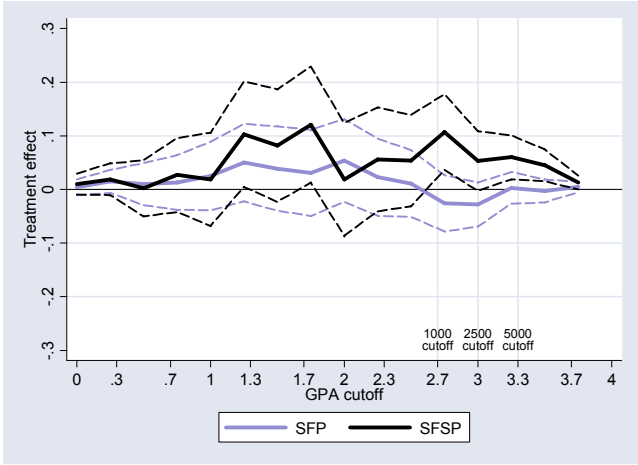


Figure 3b: Treatment Effects by Standardized GPA Cutoff, Girls

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



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



Notes: Each panel plots the coefficients on the treatment group indicators from a regression of **1**(standardized GPA≥x) on treatment and controls for high school grade quartile, mother tongue and number of credits enrolled. Standardized GPA is GPA-.3 for those in the 3rd quartile of high school grades, GPA those in the 2nd quartile, and GPA+.3 for those in the 1st quartile. Base sample is all enrolled UTM STAR students. Heteroskedasticty-robust 90% confidence intervals are indicated with dashed lines.

Table 8: Interaction Effects for First-year Grade

	All Students			Students with ≥4 Courses		
Program	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (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	-1.266	0.548	0.209	-1.524	1.428
	[1.701]	[2.737]	[2.219]	[1.736]	[2.909]	[2.198]
× low HS GPA	0.023	3.845	-2.639	-0.102	2.293	-1.668
	[1.764]	[3.054]	[2.069]	[1.768]	[3.076]	[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	-2.063	1.657	0.425	0.593	0.339
	[1.834]	[3.063]	[2.254]	[1.818]	[3.137]	[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	-2.710	-8.374	-0.079	-3.77	-9.62	-1.01
	[1.602]*	[2.873]***	[1.876]	[1.711]**	[3.144]***	[1.991]
× low HS GPA	3.222	7.400	1.623	3.31	6.20	2.81
	[1.713]*	[2.749]***	[2.202]	[1.817]*	[3.069]**	[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	1.406	4.430	0.003	2.35	5.78	0.47
	[1.838]	[3.274]	[2.202]	[1.963]	[3.653]	[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	-1.351	3.173	0.108	-4.596	2.890
	[2.555]	[4.150]	[3.273]	[2.678]	[4.708]	[3.226]
× low HS GPA	3.285	2.144	4.294	3.942	3.673	5.059
	[2.530]	[4.140]	[3.267]	[2.469]	[4.018]	[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	-1.336	1.033	-4.024	0.595	5.660	-3.109
	[2.837]	[4.684]	[3.306]	[2.945]	[5.054]	[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 University of Toronto at Mississauga (UTM) 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 9: Treatment Effects on First-year Grade, 2SLS Estimates

	Unrestricted			SSP/Any SFP		
Program	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (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	-1.598 [1.760]	-2.74 [3.509]	-0.434 [1.868]	-1.620 [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]	1.111 [1.111]			
SFSP Participant	1.778 [1.551]	0.304 [2.506]	2.599 [1.970]			
Observations	1561	661	900	1561	661	900
			Panel B: 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 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]	-1.108 [1.472]	1.488 [1.065]
SFP Participant	-0.435 [0.955]	-2.449 [1.657]	0.767 [1.166]			
SFSP Participant	2.912 [1.481]**	2.123 [2.441]	3.532 [1.877]*			
Observations	1345	554	791	1345	554	791
			Panel C: Students with Fall Grades			
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	0.305 [1.826]	-2.34 [3.605]	2.48 [1.867]	0.269 [1.825]	-2.37 [3.603]	2.45 [1.864]
Any SFP Participant				1.579 [0.847]*	0.343 [1.454]	2.48 [1.034]**
SFP Participant	1.033 [0.918]	-0.160 [1.576]	1.881 [1.128]*			
SFSP Participant	2.965 [1.509]**	1.417 [2.576]	4.192 [1.820]**			
Observations	1397	602	795	1397	602	795

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 University of Toronto at Mississauga (UTM) 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%

Appendix: Interaction Effects for Fall Grade and First-year Grade, Fall Grades Sample

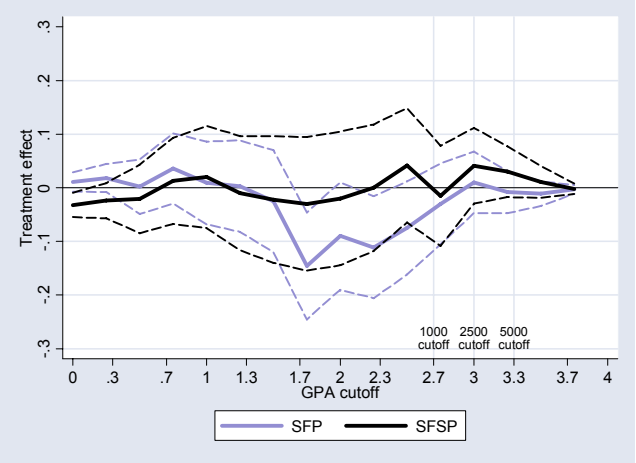
	Fall Grade			First-year Grade		
Program	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (6)
Control mean	63.8 (12.1)	65.1 (11.9)	62.8 (12.1)	60.2 (12.9)	61.0 (13.4)	59.6 (12.4)
SSP	0.036 [1.599]	-1.430 [2.635]	0.990 [1.999]	-0.260 [1.778]	-1.097 [2.992]	0.362 [2.239]
× low HS GPA	2.208 [1.800]	5.676 [2.860]**	-0.032 [2.285]	1.808 [1.822]	4.565 [3.182]	0.119 [2.031]
Total effect	2.244 [1.800]	4.246 [2.732]	0.958 [2.412]	1.548 [1.796]	3.468 [3.165]	0.481 [2.154]
× work≥5 hrs/wk	-0.888 [1.794]	-2.327 [2.869]	0.824 [2.285]	0.290 [1.899]	-2.632 [3.275]	2.886 [2.272]
Total effect	1.357 [1.616]	1.919 [2.546]	1.782 [2.056]	1.837 [1.528]	0.836 [2.655]	3.367 [1.700]
SFP	-0.226 [1.723]	-2.790 [2.380]	0.982 [2.266]	-2.18 [1.552]	-6.33 [2.595]**	-0.17 [1.905]
× low HS GPA	0.646 [1.876]	1.771 [2.667]	-0.111 [2.533]	0.92 [1.757]	2.00 [2.737]	0.54 [2.273]
Total effect	3.421 [1.958]	2.799 [2.768]	4.429 [2.785]	2.595 [1.813]	2.708 [2.524]	2.960 [2.558]
× work≥5 hrs/wk	1.406 [1.838]	4.430 [3.274]	0.003 [2.202]	2.35 [1.963]	5.78 [3.653]	0.47 [2.325]
Total effect	4.067 [1.489]	4.569 [1.856]	4.317 [2.183]	3.519 [1.472]	4.704 [2.146]	3.500 [2.020]
SFSP	2.553 [2.130]	1.901 [4.133]	4.278 [2.522]*	0.943 [2.662]	-1.099 [4.764]	3.553 [3.258]
× low HS GPA	2.143 [2.472]	1.924 [3.960]	1.292 [3.143]	3.555 [2.496]	3.273 [4.122]	3.668 [2.947]
Total effect	4.696 [2.621]	3.825 [4.279]	5.570 [3.353]	4.498 [3.078]	2.174 [5.340]	7.221 [3.194]
× work≥5 hrs/wk	-1.302 [2.566]	-2.637 [4.423]	-1.458 [3.252]	0.274 [2.964]	2.002 [5.111]	-2.229 [3.329]
Total effect	3.394 [2.194]	1.188 [3.526]	4.112 [2.760]	4.772 [1.970]	4.175 [3.287]	4.992 [2.424]
Observations	1273	539	734	1273	539	734

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 University of Toronto at Mississauga (UTM) students participating in the STAR program with fall grades 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.

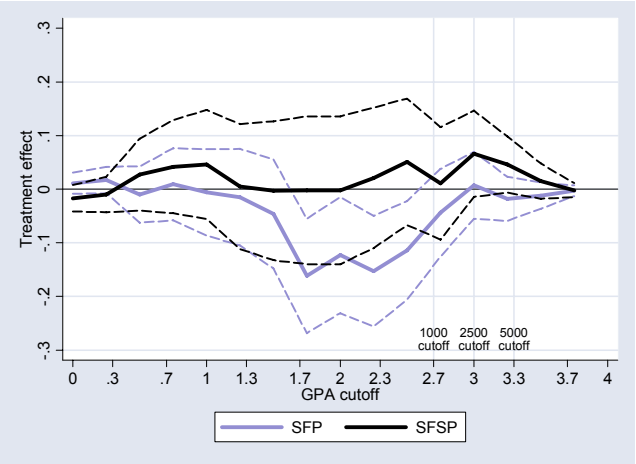
* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix Figure A: Treatment Effects by Standardized GPA Cutoff, Boys with Fall Grades

Panel A: SFP vs. control and SFSP vs. control, Fall grades sample

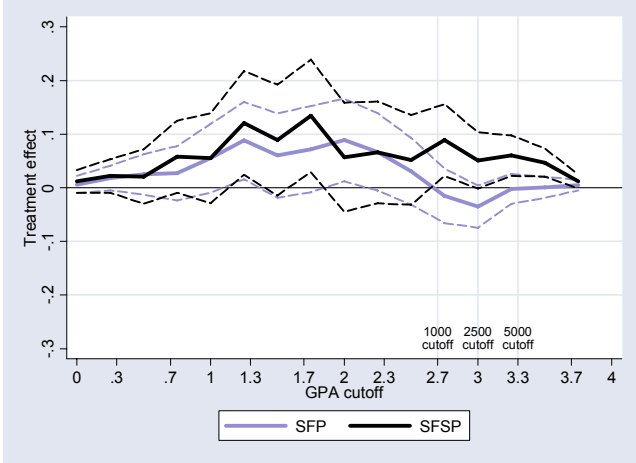


Panel B: SFP vs. control and SFSP vs. control, Fall grades and ≥4 courses sample

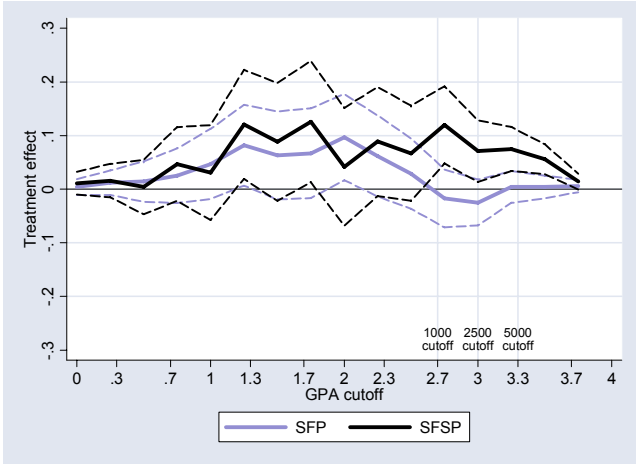


Appendix Figure B: Treatment Effects by Standardized GPA Cutoff, Girls with Fall Grades

Panel A: SFP vs. control and SFSP vs. control, Fall grades sample



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



Notes: Each panel plots the coefficients on the treatment group indicators from a regression of **1**(standardized GPA≥x) on treatment and controls for high school grade quartile, mother tongue and number of credits enrolled. Standardized GPA is GPA-.3 for those in the 3rd quartile of high school grades, GPA those in the 2nd quartile, and GPA+.3 for those in the 1st quartile. Base sample is UTM STAR students with at least one fall grade. Heteroskedasticty-robust 90% confidence intervals are indicated with dashed lines.