Demonstrating selection effects in evaluation of a first year seminar.

Kevin Laughren[a] * , Lara Aknin[a], Dai Heide[a], Panayiotis Pappas[a], and Milan Singh[a]

[a]Simon Fraser University, Burnaby, BC, Canada

ARTICLE HISTORY
Compiled July 10, 2019

ABSTRACT
We examine the effect of a first year seminar on student retention, academic status, and grades at a Canadian public university using a novel control group. Using a difference-in-differences framework we demonstrate that students who enroll in a seminar have significantly higher GPA when compared to a matched sample or the entire cohort, but that these effects disappear when we control for 'willingness to enroll'. We suggest that studies failing to control for students' (unobserved) willingness to enroll in a seminar may be subject to greater selection bias.

KEYWORDS
Program Evaluation; First Year Experience; Seminar; Difference-in-Differences; Higher Education; Retention

1. Opening Paragraph

Existing studies that set out to measure the causal effect of educational programs on student outcomes are often limited to designs other than the randomized control trial (RCT) because research ethics require that a student opt-in to such programs, even when eligibility is randomly assigned. The result is that analysis of student outcomes cannot rely on random assignment to believe that the students participating in the intervention are comparable to those who do not participate. This means that any differences between treatment and control groups that we observe in such analyses could be caused not by the treatment itself, but by unobserved and meaningful differences in the sample of students who select into such programs relative to the rest of their cohort. We refer to significant effects driven by this type of sample selection as *selection effects*. Several experimental design paradigms attempt to alleviate these concerns, including studies that generate a control group by selecting a group of students who match the treated students on observable characteristics such as grades, gender, race, and socioeconomic status. We study the causal effects of a first-year seminar program at a public university using a difference-in-differences framework (Wooldridge, 2010) and find that seminar enrollment has a positive causal effect on GPA when using the entire cohort or a matched sample as the control group. Our linear probability and logistic binary regressions using these same control groups suggest that seminar enrollment positively affects the probability of first year student retention into their second year. Our contribution to the literature is to elicit students' *willingness to enroll* prior to

* CONTACT Email:klaughre@sfu.ca

the educational intervention, which we believe is a meaningful and unobserved covariate in most non-RCT studies of educational programs. When we include this control variable our causal effects are no longer statistically significant, suggesting that the causal effects we observe with our typical control groups are actually selection effects. Many studies of educational interventions risk confounding genuine causal effects with underlying differences between students who opt-in to treatment versus those who do not.

2. Introduction

Student satisfaction and retention have been a focus of undergraduate administrators for decades, and along with GPA, have been the outcomes of interest for a large program evaluation literature in higher education research. Evaluation of these programs dates at least as far back as Pascarella and Terenzini (1979). Pascarella and Terenzini (1991) provides a survey of early studies, and Bailey (2005) provides a survey of interventions specific to community colleges where low retention is even more pernicious than at degree-granting institutions.

Aside from a few older studies such as Strumpf and Hunt (1993), research in this literature has not been able to conduct experiments following the Randomized Control Trial (RCT) methodology due to ethical concerns with forcing students into any educational intervention, whether it be a specific course or an academic support program. Instead, students select themselves into such programs, and so we cannot rely on random assignment to believe that the students participating in the program are comparable to students who do not participate. This means that any differences between treatment and control groups that we observe in such analyses could be caused not by the treatment itself, but by unobserved and meaningful differences in the sample of students who select into such programs relative to the rest of their cohort. We refer to significant effects driven by this type of sample selection as *selection effects*.

Our contribution to the educational program evaluation literature is to collect a novel proxy for a student's willingness to enroll in a first year seminar (FYS) and demonstrate that results drawn from comparisons to an entire cohort or matched sample may be driven by selection effects, as the results are not robust when comparing to our novel control group.

We believe collecting such a proxy provides a superior control to a matched pairs protocol, because generating unbiased estimates with a matched control group requires a *selection on observables* assumption (Wooldridge, 2010). This assumption states that after we control for the observable characteristics on which a match is generated - demographics and high school grades - two students of the same observable type are equally likely to select into a seminar. However if this assumption is not satisfied, the estimates generated by this protocol are biased.

Considering that FYS courses require substantially different types of work from students - in particular frequent writing and speaking in class, with little or no weight on final exams - we think it is plausible that the selection on observables assumption is not satisfied in this instance. Demographics and high school grades may not be enough information to accurately predict the probability that a student enrolls themself in a seminar. A student's desire to be evaluated on their writing and speaking in a seminar relative to their desire to be evaluated on exam performance in a lecture class is an unobserved variable that, when omitted from regression, can introduce bias to estimates of the effects of seminar participation.

To proxy for this unobserved variable, we sent an incentivized email survey to all incoming domestic Faculty of Arts and Social Sciences (Faculty) students at a large public Canadian university prior to the Fall 2017 enrollment period. This survey included detailed information on the seminar program overall (small classes, award winning faculty, emphasis on research, writing, and speaking) as well as titles and synopses of the ten seminars being offered. Students were asked to provide a Yes/No answer to whether they were willing to enroll in each of the ten, and told their responses could affect whether a seminar would run in Fall 2017. 219 eligible students responded to this survey.[1] We use those students who did not actually enroll in a seminar but indicated 'Yes' to being willing to enroll in some seminar as our preferred control group relative to the students who did enroll in a seminar.

In comparisons of seminar students to the entire cohort or a matched sample, we find that the first year cumulative GPA of seminar students is significantly greater. However in our preferred model specifications that control for willingness to enroll, we find enrollment in a first year seminar provides no significant benefit on student well-being, grades, or continuation to second year. We believe this is evidence that selection effects may be driving the significant results in program evaluation papers that implicitly rely on a selection on observables assumption.

2.1. Related Literature

Broadly, this is a program evaluation study, with methods informed by the Wooldridge (2010) text on econometrics for cross-section and panel data. Some recent studies in higher education have used similar program evaluation techniques including regression discontinuity Moss and Yeaton (2013), difference-in-differences, and propensity score matching Sneyers and Witte (2017).

There is a substantial literature with a specific focus on college student retention as a response to an educational intervention using linear probability models (OLS): (Andrade, 2009; Crissman, 2001; Sommet, Quiamzade, Jury, & Mugny, 2015; Swanson, Vaughan, & Wilkinson, 2017; Webster & Showers, 2011), or logistic binary variables regression Venuleo, Mossi, and Salvatore (2016); including several with a particular focus on minorities or international students: (Andrade, 2009; Barlow & Villarejo, 2004). The *What Works Clearinghouse* is an online repository of results from educational program evaluation studies with strict criteria for inclusion, and includes a section specific to post-secondary interventions. [2]

Shanley and Witten (1990) and Fidler (1991) conducted a 15-year study at University of South Carolina comparing seminar students to all other first years, and found significant improvement in retention that could not be explained by observed demographic and high school variables.

Many of the above papers cite Tinto (1975), Astin (1993), or Seidman (2005) as sources of theoretical hypotheses that the types of activities in a first year seminar could improve retention. We do not explicitly use these models to generate hypotheses but rely on reduced-form models to test intuitive hypotheses that are in line with these models.

The vast majority of studies - including our own - are cases where students opt into treatment, whether it be a seminar, mentor program, math prep course, etc. One exception is Strumpf and Hunt (1993) in which the researches collected motivation to enroll and randomly assigned a subset of the motivated students to a seminar, leaving the rest as a control group. They found significant effects on retention and academic

standing. (Random assignment of courses would not be feasible or pass ethics approval in most public universities today). Angrist, Lang, and Oreopoulos (2009) randomly assign *eligibility*, but students had to actively consent to receive academic support services (with 55% consent) or financial GPA incentives (with 87% consent).

Meta analyses of retention studies (Colton, Ulysses J. Connor, Shultz, & Easter, 1999; Fike & Fike, 2008; Fong et al., 2017) focus on demographic predictors of retention, reach little consensus, and point out methodological challenges. Clark and Cundiff (2011); Reid, Reynolds, and Perkins-Auman (2014) are surveys that point out methodological issues in previous studies, particularly the inability of previous causal analyses to rule out potential confounding explanations, in particular unobserved characteristics affecting selection.

Many studies relied on creating a group of control students through matching, i.e., ensuring that the control group had a student who was similar to a student in the seminar based on recorded characteristics such as gender, ethnicity, and high school grades (Campbell & Campbell, 1997; Hendel, 2007; Miller & Lesik, 2014; Schnell & Doetkott, 2003). These analyses implicitly rely on a selection on observables assumption (Wooldridge, 2010), which essentially requires that there is no difference in error distributions after controlling for observable variables. But they are still omitting a known difference between the two groups willingness to enroll in a seminar-style course over a traditional large lecture course.

3. Methods

The Faculty launched nine[3] first year seminars led by highly regarded permanent faculty members in the Fall 2017 semester. These seminars were open only to domestic first-year non-transfer students in the Faculty. Each seminar enrollment was limited to 25 students.

The rationale for offering the seminars was twofold. First, to determine whether offering students this kind of first-year experience would positively impact their GPA and their overall well-being in addition to reducing the likelihood that they will leave the university without graduating. Second, such seminars were thought to allow students to develop closer bonds with faculty and to feel more closely connected with SFU and with their peers at an early stage in their academic careers.

The pilot was meant to examine the long-term effects of participation in a seminar on GPA and student retention. While meaningful data on graduation will not be available for several years, we report on one-year retention and first year GPA. To measure this, we identified a control group by surveying all incoming domestic Faculty students concerning their level of interest in the courses (regardless of whether they in fact enroll). Focusing only on those students willing to enroll in such a seminar helps us to rule out selection effects, which are inevitable in this setting where we cannot choose which students enroll in a seminar. In our causal effects framework, we compare first year domestic students who enroll in FYS with willing students who did not enroll, and determine whether participation in FYS positively affects GPA, academic status, and retention.

If nothing else, our willingness proxy identifies students who are reading and actively responding to emails prior to and early in the enrollment period, which is an unobserved but meaningful covariate in the majority of other educational program evaluation studies.

This project also examines the immediate impact of participation in FYS on student

engagement and well-being. We collected validated measures of well-being in surveys of students at two times: prior to the start of the semester, and at the end of the semester. We use a difference-in-differences framework to examine whether participation in FYS positively impacts student well-being and engagement, while controlling for student selection, beginning versus end of term effects, and covariates such as entering (high school) GPA, gender, and course load.

In total, 124 domestic first year students enrolled in one of the seminars. These students achieved a 2.57 GPA in their non-FYS courses in Fall 2017, compared to a 2.31 GPA in non-FYS courses for the entire cohort. Of course there are selection issues that inhibit us from making a causal interpretation between these two numbers.

3.1. Data

3.1.1. University Registrar

The university's registrar provided the majority of our data, including all data on demographics, enrollment, academic status, and grades. Most of our analyses require only these data, except where we controlled for willingness to enroll or wanted to look for effects of FYS on psychological measures of well-being.

3.1.2. Willingness to enroll

We sent an incentivized email survey to all incoming domestic Faculty students prior to the Fall 2017 enrollment period. This survey included detailed information on the seminar program overall (small classes, award winning instructors, emphasis on research, writing, and speaking) as well as titles and synopses of the ten seminars being offered. Students were asked to provide a Yes/No answer to whether they were willing to enroll in each of the ten, and told their responses could affect whether a seminar would run in Fall 2017. 219 eligible students responded to this survey. We use those students who did not actually enroll in a seminar but indicated Yes to being willing to enroll in some seminar as our control group relative to the students who did enroll in a seminar.

3.1.3. Well-being and engagement

Our surveys of well-being and engagement were conducted online, with domestic first-year Faculty students recruited directly via email. Completion was incentivized with a draw for Visa gift cards in $50CAD and $100CAD denominations. Individual responses to a series of Likert-scale style questions are aggregated into individual measures of Depression (CESD) (Radloff, 1977), Loneliness (D. Russell, Peplau, & Ferguson, 1978), Satisfaction with Life (Diener, Emmons, Larsen, & Griffin, 1985), Social Connection (Lee, Draper, & Lee, 2001), Anxiety (Spitzer, Kroenke, Williams, & Löwe, 2006), Flourishing, and frequency of Positive and Negative Emotions (SPANE) (Diener et al., 2010). The measures were chosen for the survey based on their common usage in psychology research and strong performance on validity measures such as test-retest reliability, internal consistency, and convergent validity, (Aishvarya et al., 2014; D. W. Russell, 1996; Silva & Caetano, 2013). We also asked about a students feelings of Belonging at the University, and about their number of close friends and acquaintances.

311 students completed at least one well-being survey, 71 of whom completed it both before and after the semester.

To examine how the measures change over the course of a semester, we restrict attention to the 71 students who completed a survey twice (to control for sample selection). Within these students, we calculate the correlation coefficient between well-being measures, as well as covariates (high school GPA, gender), treatment status (TREATMENT), and a dummy variable for measurements done post-seminar (POST). In Appendix A, we provide results of a t-test on each Pearson correlation coefficient to determine if the measured correlations are significantly different from zero. None of the measures of well-being are significantly correlated with POST, which tells us that the overall average response of these students did not change significantly over the Fall 2017 semester. The correlation matrix also suggests that female students have significantly higher measures of Negative Emotions, Anxiety, Social Connection, and Symptoms of Depression. Notably, high school grades (eGPA) are significantly correlated with Satisfaction with Life[4].

3.2. Empirical Frameworks

3.2.1. Difference-in-differences

We implement a difference-in-differences causal effects framework (Wooldridge, 2010) using ordinary least squares (OLS) and fixed effects (FE) frameworks in an attempt to identify causal effects of the seminar on well-being and GPA. The difference-in-differences framework involves running a regression of the form:

$$y_{i,t} = \alpha_i + \beta_1 * TREATMENT + \beta_2 * POST + \delta * TREATMENT : POST + \theta * X_i + u_{i,t}$$

Where $y_{i,t}$ is the outcome variable y (e.g., GPA, academic status, anxiety) for individual i measured at time t. X_i is a vector of individual i's covariate characteristics such as their high school grades, home province, gender, etc. In the OLS specification we can estimate θ, the effect of covariates on outcomes like GPA because of the assumption that all individuals have the same intercept: $\alpha_i = \alpha \quad \forall i$. In the FE specification, the coefficient on any covariates X_i that do not change across t (such as gender) are all collapsed into an individual estimate of α_i. TREATMENT = 1 for individuals who enrolled in FYS and zero otherwise. POST = 1 for observations made post-seminar (i.e., December 2017 or later) and zero otherwise. The causal effect of seminar enrollment on outcome y is measured by δ, which can be thought of as the interaction effect on TREATMENT*POST. Letting $t \in \{PRE, POST\}$, algebraic manipulation of the regression equation reveals why δ is known as the difference-in-differences estimator:

$$\delta = (\overline{y_{T,POST}} - \overline{y_{T,PRE}}) - (\overline{y_{C,POST}} - \overline{y_{C,PRE}})$$

Where T represents treatment group, C represents the control group, and the overbar represents the average over the group in the specified time period (PRE or POST). We could calculate δ by hand, but by using regression we obtain standard errors that help us understand whether the estimated value is significantly different from zero.

In all specifications, individuals who enrolled in a seminar in Fall 2017 are identified as the Treatment group, and we vary the control group to demonstrate the lack of robustness of effects to our measure of willingness to enroll. Specifically, each regression table is subdivided into three sections: where the control group are students who responded they were willing to enroll in a seminar but did not, where the control

group are students matched to a treatment student on the basis of demographics and high school grades, and finally where the control group is the entire cohort of first year domestic students.

3.2.2. Non-panel frameworks

While we have measures of well-being and grades prior to seminar enrollment, a number of measures of interest are only observed once (after the seminar) and so a difference-in-differences framework is not applicable. In particular, for grades, academic status, and enrollment we specify a linear model of the form:

$$y_{i,t} = \alpha + \beta_1 * TREATMENT + \theta * X_i + u_{i,t}$$

For binary variables (enrollment and academic status) we also estimate a logit model of the form:

$$Prob(y_{i,t} = 1) = \frac{e^{\alpha + \beta_1 * TREATMENT_i + \theta * X_i}}{1 + e^{\alpha + \beta_1 * TREATMENT_i + \theta * X_i}}$$

$$log\left(\frac{Prob(y_{i,t} = 1)}{1 - Prob(y_{i,t} = 1)}\right) = \alpha + \beta_1 * TREATMENT_i + \theta * X_i$$

4. Results

4.1. Baseline Equivalence

In this quasi-experimental methodology, it is important to establish that the treatment and control groups are comparable on observable baseline dimensions. We compare the students enrolled in a seminar in the left column to three potential control groups in Table 1. By construction, the *Match* control group is almost exactly the same as the *Seminar* group on observable dimensions. The *Willing* control group (those who took the pre-enrollment survey and indicated they were willing to enroll in a seminar, but did not end up doing so) have similar proportion of males and course load to the seminar group, but have higher high school grades (86.6 versus 84.8) and fewer students identifying as First Nations (0.7% versus 2.4% for the treatment group). These differences are not significant at a 5% level using a Mann-Whitney U test of means. Using the entire cohort as a control group introduces substantial difference in gender ratio (35.7% male versus 25.8% in the treatment group), as female students had a higher propensity to select into FYS. A number of studies have demonstrated that female students achieve higher post-secondary grades, so we should expect a comparison between *Seminar* students and the *Cohort* control group to show greater grades for the *Seminar* group.

Table 1. Baseline measures

	Seminar Group	*Control Groups*		
	TREATMENT = 1	*Willing*	*Match*	*Cohort*
Number of Students	124	151	124	1014
HighSchoolGPA (out of 100)	84.82	86.61	84.87	85.43
% Male	25.8%	25.2%	25.8%	35.6%
% First Nations	2.4%	0.7%	2.4%	1.3%
Fall2017 Course Units	11.19	11.20	11.25	11.05

Note: *p<0.1; **p<0.05; ***p<0.01

4.2. Seminar effect on GPA

Recall that the causal effect of FYS on GPA in the difference-in-differences framework, δ is the coefficient on TREATMENT:POST. Table 2 demonstrates that the estimate of this effect is positive and significant when comparing FYS students to the *Match* or *Cohort* control groups, but there is zero effect of FYS when we compare treated students to the *Willing* control group. This is the first evidence we provide of selection effects in frameworks that do not control for willingness to enroll.

Table 2. Seminar effect on GPA

	Dependent variable: GPA_Proxy					
Control Group:	*Willing*		*Match*		*Cohort*	
	OLS	*FE*	*OLS*	*FE*	*OLS*	*FE*
	(1)	(2)	(3)	(4)	(5)	(6)
TREATMENT	−0.099		−0.022		−0.071	
	(0.077)		(0.089)		(0.062)	
POST	−1.262***	−1.264***	−1.549***	−1.549***	−1.481***	−1.482***
	(0.073)	(0.060)	(0.089)	(0.080)	(0.029)	(0.025)
GENDERM	−0.280***		−0.395***		−0.210***	
	(0.062)		(0.072)		(0.029)	
TREATMENT:POST	−0.0001	0.002	0.286**	0.286**	0.219**	0.219***
	(0.109)	(0.089)	(0.126)	(0.113)	(0.088)	(0.077)
Constant	3.989***		3.941***		3.943***	
	(0.054)		(0.066)		(0.023)	
Observations	550	550	496	496	2,276	2,276
R^2	0.510	0.749	0.521	0.717	0.559	0.767

Note: *p<0.1; **p<0.05; ***p<0.01

4.3. Seminar effect on well-being

We conducted two optional surveys prior to and following the seminar in Fall 2017 to collect psychological measures of well-being. If the seminar had significant effects on academic outcomes, these results might help us to identify a mechanism by which this happens (e.g., by increasing feelings of belonging to the school community). We used the same difference-in-differences framework as we did for GPA in an attempt to measure any causal effects of the seminar on these outcomes. The sample is smaller than the GPA analysis however, because we are limited to those students who chose to complete both surveys of well-being, and who identified themselves as willing seminar enrollees either through our pre-enrollment survey or by enrolling themselves. The table below provides our difference-in-differences fixed effects estimates with each column representing a different measure of well-being.

Table 3 shows that of the nine psychological measures, we see a significant causal effect of the seminar on only one: CESD (a measure of how frequently an individual experiences symptoms of depression). This effect is significant only to a 10% significance level. The probability of observing at least one significant result at a 10% level when conducting nine independent tests is substantial, so we interpret these results as inconclusive across the board. There seems to be no causal effect of enrolling in a first year seminar on any of the following: symptoms of depression, loneliness, flourishing, anxiety, satisfaction with life, social connection, frequency of positive emotions, frequency of negative emotions, or belonging to the school community.

Table 3. Seminar Effect on Survey Measures of Well-being

	Dependent variable:								
	Symptoms of Depression	Loneliness	Flourishing	Anxiety	Satisfaction with Life	Social Connection	Frequency of Pos. Emotions	Frequency of Neg. Emotions	Belonging at University
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
POST	0.201	0.625	0.005	0.783	−0.192	0.059	0.080	0.680	−0.522**
	(0.127)	(0.633)	(0.150)	(0.973)	(0.216)	(0.049)	(0.693)	(0.836)	(0.225)
TREATMENT:POST	−0.373*	−0.625	−0.218	−1.253	0.251	−0.099	0.861	0.555	0.584
	(0.198)	(1.001)	(0.233)	(1.493)	(0.340)	(0.077)	(1.090)	(1.314)	(0.351)
Observations	82	83	84	83	85	80	85	85	82
R^2	0.093	0.025	0.036	0.021	0.021	0.052	0.031	0.051	0.128

Note: Observation numbers vary slightly as all questions were optional

*p<0.1; **p<0.05; ***p<0.01

4.4. Seminar effect on academic status and retention

Tables 4 and 5 provide a linear probability model (OLS) and a logistic binary variable regression on Spring 2018 academic standing and Fall 2018 (second year) enrollment, respectively.

Table 4 shows that FYS enrollment has a positive effect on academic standing when comparing enrolled students to the *Match* or *Cohort* control groups, but again there is no significant effect when FYS students are compared to the *Willing* control group.

Table 5 shows that FYS students are significantly more likely to be retained to second year (as measured by Fall 2018 enrollment) than the control group generated by matching covariates, but that there is no significant effect of FYS on enrollment relative to the *Cohort* or *Willing* control groups.

Table 4. Seminar effect on academic standing after two semesters

	Dependent variable: Spring 2018 Good Academic Standing					
Control Group:	*Willing*		*Match*		*Cohort*	
	OLS	*logistic*	*OLS*	*logistic*	*OLS*	*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)
SEMINARTREATMENT	0.040	0.386	0.121**	0.726**	0.090**	0.561**
	(0.047)	(0.332)	(0.053)	(0.313)	(0.041)	(0.248)
HIGHSCHOOLGPA	0.017***	0.192***	0.008**	0.062**	0.016***	0.127***
	(0.003)	(0.038)	(0.003)	(0.030)	(0.002)	(0.016)
GENDERM	−0.101*	−0.606*	−0.223***	−1.099***	−0.066**	−0.314**
	(0.054)	(0.340)	(0.062)	(0.325)	(0.027)	(0.142)
Constant	−0.626**	−14.899***	0.091	−4.200	−0.634***	−9.785***
	(0.280)	(3.243)	(0.257)	(2.595)	(0.172)	(1.345)
Observations	275	275	248	248	1,138	1,138
R^2	0.118		0.109		0.069	
Log Likelihood		−119.284		−128.981		−624.536

Note: Logistic coefficients (not marginal effects) are displayed so they can be compared to their standard errors
*p<0.1; **p<0.05; ***p<0.01

Table 5. Seminar effect on second-year enrollment

	Dependent variable: Fall 2018 Enrollment					
Control Group:	*Willing*		*Match*		*Cohort*	
	OLS	*logistic*	*OLS*	*logistic*	*OLS*	*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)
SEMINARTREATMENT	−0.018	−0.154	0.113**	0.704**	0.052	0.366
	(0.042)	(0.357)	(0.051)	(0.324)	(0.038)	(0.264)
HIGHSCHOOLGPA	0.005*	0.032	0.001	0.003	0.005**	0.025**
	(0.003)	(0.022)	(0.003)	(0.016)	(0.002)	(0.011)
GENDERM	−0.043	−0.351	−0.094	−0.543	0.005	0.030
	(0.048)	(0.386)	(0.060)	(0.349)	(0.025)	(0.159)
Constant	0.435*	−0.702	0.709***	0.888	0.408***	−0.750
	(0.252)	(1.941)	(0.248)	(1.386)	(0.157)	(0.920)
Observations	275	275	248	248	1,138	1,138
R^2	0.019		0.030		0.007	
Log Likelihood		−108.269		−123.625		−560.751

Note: Logistic coefficients (not marginal effects) are displayed so they can be compared to their standard errors
*p<0.1; **p<0.05; ***p<0.01

5. Discussion

We are able to replicate significant positive effects of a first year seminar (FYS) on GPA, academic status, and retention when we compare participants to convenient control groups used in previous studies, whether they be a group matched on demographics, or a comparison to the overall cohort. However, when we use our pre-enrollment survey measure to proxy for willingness to enroll, and compare FYS students only to others who are interested in such a course but do not enroll, all significant effects disappear. We see in this evidence that some previous positive program evaluation results on FYS may have been driven by selection effects. Controlling only for observable characteristics such as demographics, high school grades, and location does not appear to be sufficient to use such groups for causal effects analysis because the selection on observables assumption is violated. Specifically, after controlling for observable characteristics, it is not true that all students are equally likely to enroll in FYS; most studies omit an unobserved characteristic *willingness to enroll* that captures student preferences for small interactive classrooms with less weight on exams. This unobserved and uncontrolled preference may be driving spurious results in previous program evaluations of first year seminars in post secondary education.

Funding

Research assistance and survey incentives were funded by Teaching and Learning Development Grants (TLDG Projects G0233, G0234) from the Institute for the Study of Teaching and Learning in the Disciplines at Simon Fraser University. This research was conducted during the corresponding author's doctoral studies, which have been partially funded by a doctoral fellowship from Social Sciences and Humanities Resource Council (SSHRC).

Notes

[1] And an additional 24 ineligible students who found the survey link through other means

[2] https://ies.ed.gov/ncee/wwc/FWW/Results?filters=,Postsecondary

[3] One of the ten seminars listed in the survey was cancelled due to lack of enrollment

[4] There are a number of significant correlations between well-being variables; for example, our measure of Belonging is significantly negatively correlated with Loneliness, Social Connection, and Symptoms of Depression.

References

Aishvarya, S., Maniam, T., Karuthan, C., Sidi, H., Jaafar, N. R. N., & Oei, T. P. S. (2014). Psychometric properties and validation of the satisfaction with life scale in psychiatric and medical outpatients in malaysia. *Comprehensive Psychiatry*, *55*, S101–S106.

Andrade, M. S. (2009). The value of a first-year seminar: International students' insights in retrospect. *Journal of College Student Retention: Research, Theory & Practice*, *10*(4), 483-506. Retrieved from `https://doi.org/10.2190/CS.10.4.e`

Angrist, J., Lang, D., & Oreopoulos, P. (2009, January). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *1*(1), 136-63. Retrieved from `http://www.aeaweb.org/articles?id=10.1257/app.1.1.136`

Astin, A. W. (1993). College retention rates are often misleading. *Chronicle of Higher Education*, *40*(5), A48–A48.

Bailey, T. R. (2005). Paths to persistence: An analysis of research on program effectiveness at community colleges.

Barlow, A. E., & Villarejo, M. (2004). Making a difference for minorities: Evaluation of an educational enrichment program. *Journal of Research in Science Teaching*, *41*(9), 861-881. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/tea.20029

Campbell, T. A., & Campbell, D. E. (1997, Dec 01). Faculty/student mentor program: Effects on academic performance and retention. *Research in Higher Education*, *38*(6), 727–742. Retrieved from https://doi.org/10.1023/A:1024911904627

Clark, M., & Cundiff, N. L. (2011). Assessing the effectiveness of a college freshman seminar using propensity score adjustments. *Research in Higher Education*, *52*(6), 616–639.

Colton, G. M., Ulysses J. Connor, J., Shultz, E. L., & Easter, L. M. (1999). Fighting attrition: One freshman year program that targets academic progress and retention for at-risk students. *Journal of College Student Retention: Research, Theory & Practice*, *1*(2), 147-162. Retrieved from https://doi.org/10.2190/FTPB-1LQ7-XBUX-J1RY

Crissman, J. L. (2001). The impact of clustering first year seminars with english composition courses on new students' retention rates. *Journal of College Student Retention: Research, Theory & Practice*, *3*(2), 137-152. Retrieved from https://doi.org/10.2190/FJHU-RT1X-GA6Y-EME5

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71–75.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-w., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*(2), 143–156.

Fidler, P. (1991). Relationship of freshman orientation seminars to sophomore return rates. *Journal of the First-Year Experience & Students in Transition*, *3*(1), 7–38.

Fike, D. S., & Fike, R. (2008). Predictors of first-year student retention in the community college. *Community College Review*, *36*(2), 68-88. Retrieved from https://doi.org/10.1177/0091552108320222

Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research*, *87*(2), 388-424. Retrieved from https://doi.org/10.3102/0034654316653479

Hendel, D. D. (2007). Efficacy of participating in a first-year seminar on student satisfaction and retention. *Journal of College Student Retention: Research, Theory & Practice*, *8*(4), 413-423. Retrieved from https://doi.org/10.2190/G5K7-3529-4X22-8236

Lee, R. M., Draper, M., & Lee, S. (2001). Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: Testing a mediator model. *Journal of counseling psychology*, *48*(3), 310.

Miller, J. W., & Lesik, S. S. (2014). College persistence over time and participation in a first-year seminar. *Journal of College Student Retention: Research, Theory & Practice*, *16*(3), 373-390. Retrieved from https://doi.org/10.2190/CS.16.3.d

Moss, B. G., & Yeaton, W. H. (2013). Evaluating effects of developmental education for college students using a regression discontinuity design. *Evaluation Review*, *37*(5), 370-404. Retrieved from https://doi.org/10.1177/0193841X14523620 (PMID: 24662603)

Pascarella, E. T., & Terenzini, P. T. (1979). Student-faculty informal contact and college persistence: A further investigation. *The Journal of Educational Research*, *72*(4), 214–218.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students* (Vol. 1991). Jossey-Bass San Francisco.

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, *1*(3), 385–401.

Reid, K. M., Reynolds, R. E., & Perkins-Auman, P. G. (2014). College first-year seminars: What are we doing, what should we be doing? *Journal of Col-*

lege Student Retention: Research, Theory & Practice, *16*(1), 73-93. Retrieved from https://doi.org/10.2190/CS.16.1.d

Russell, D., Peplau, L. A., & Ferguson, M. L. (1978). Developing a measure of loneliness. *Journal of personality assessment*, *42*(3), 290–294.

Russell, D. W. (1996). Ucla loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of personality assessment*, *66*(1), 20–40.

Schnell, C. A., & Doetkott, C. D. (2003). First year seminars produce long-term impact. *Journal of College Student Retention: Research, Theory & Practice*, *4*(4), 377-391. Retrieved from https://doi.org/10.2190/NKPN-8B33-V7CY-L7W1

Seidman, A. (2005). Minority student retention: Resources for practitioners. *New directions for institutional research*, *2005*(125), 7–24.

Shanley, M. G., & Witten, C. H. (1990). University 101 freshman seminar course: A longitudinal study of persistence, retention, and graduation rates. *NASPA Journal*, *27*(4), 344–352.

Silva, A. J., & Caetano, A. (2013). Validation of the flourishing scale and scale of positive and negative experience in portugal. *Social Indicators Research*, *110*(2), 469–478.

Sneyers, E., & Witte, K. D. (2017). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction. evidence from the netherlands. *Studies in Higher Education*, *42*(2), 354-389. Retrieved from https://doi.org/10.1080/03075079.2015.1049143

Sommet, N., Quiamzade, A., Jury, M., & Mugny, G. (2015). The student-institution fit at university: interactive effects of academic competition and social class on achievement goals. *Frontiers in Psychology*, *6*, 769. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2015.00769

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the gad-7. *Archives of internal medicine*, *166*(10), 1092–1097.

Strumpf, G., & Hunt, P. (1993). The effects of an orientation course on the retention and academic standing of entering freshmen, controlling for the volunteer effect. *Journal of The First-Year Experience and Students in Transition*, *5*(1), 7–14.

Swanson, N. M., Vaughan, A. L., & Wilkinson, B. D. (2017). First-year seminars: Supporting male college students long-term academic success. *Journal of College Student Retention: Research, Theory & Practice*, *18*(4), 386-400. Retrieved from https://doi.org/10.1177/1521025115604811

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, *45*(1), 89–125.

Venuleo, C., Mossi, P., & Salvatore, S. (2016). Educational subculture and dropping out in higher education: a longitudinal case study. *Studies in Higher Education*, *41*(2), 321-342. Retrieved from https://doi.org/10.1080/03075079.2014.927847

Webster, A. L., & Showers, V. E. (2011). Measuring predictors of student retention rates. *American Journal of Economics and Business Administration*, *3*(2), 301–311.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

6. Appendices

Appendix A. Correlation of Well-Being Measures and Covariates

Correlation coefficients

	TREATMENT	POST	Loneliness	SPANE_Pos	SPANE_Neg	SWLS	Flourishing	Anxiety	SocialConnection	CESD	BelongSFU	eGPA	FEMALE
TREATMENT	1.00	0.00	0.02	0.12	0.01	-0.04	0.07	0.02	-0.08	0.04	-0.12	-0.23	-0.16
POST	0.00	1.00	0.04	0.05	0.11	-0.08	-0.11	-0.01	0.03	-0.01	-0.09	-0.02	0.04
Loneliness	0.02	0.04	1.00	-0.19	0.19	-0.29	-0.39	0.09	0.38	0.19	-0.44	-0.01	0.09
SPANE_Pos	0.12	0.05	-0.19	1.00	-0.38	0.51	0.53	-0.42	-0.30	-0.33	0.10	0.11	0.06
SPANE_Neg	0.01	0.11	0.19	-0.38	1.00	-0.36	-0.23	0.65	0.29	0.63	-0.14	0.00	0.31
SWLS	-0.04	-0.08	-0.29	0.51	-0.36	1.00	0.62	-0.38	-0.40	-0.40	0.15	0.35	-0.06
Flourishing	0.07	-0.11	-0.39	0.53	-0.23	0.62	1.00	0.00	-0.50	-0.06	0.19	0.01	0.18
Anxiety	0.02	-0.01	0.09	-0.42	0.65	-0.38	0.00	1.00	0.26	0.80	-0.10	-0.06	0.34
SocialConnection	-0.08	0.03	0.38	-0.30	0.29	-0.40	-0.50	0.26	1.00	0.32	-0.34	0.14	0.31
CESD	0.04	-0.01	0.19	-0.33	0.63	-0.40	-0.06	0.80	0.32	1.00	-0.33	-0.10	0.38
BelongSFU	-0.12	-0.09	-0.44	0.10	-0.14	0.15	0.19	-0.10	-0.34	-0.33	1.00	0.14	-0.15
eGPA	-0.23	-0.02	-0.01	0.11	0.00	0.35	0.01	-0.06	0.14	-0.10	0.14	1.00	0.32
FEMALE	-0.16	0.04	0.09	0.06	0.31	-0.06	0.18	0.34	0.31	0.38	-0.15	0.32	1.00

P-values of test: is correlation coefficient = 0 ?

P

	TREATMENT	POST	Loneliness	SPANE_Pos	SPANE_Neg	SWLS	Flourishing	Anxiety	SocialConnection	CESD	BelongSFU	eGPA	FEMALE
TREATMENT		1.0000	0.8492	0.3332	0.9111	0.7391	0.5389	0.8621	0.5090	0.7217	0.2972	0.0559	0.1754
POST	1.0000		0.7451	0.6607	0.3427	0.5169	0.3764	0.9511	0.7842	0.9349	0.4537	0.8694	0.7633
Loneliness	0.8492	0.7451		0.1170	0.1013	0.0135	0.0008	0.4606	0.0009	0.1047	0.0001	0.9116	0.4382
SPANE_Pos	0.3332	0.6607	0.1170		0.0009	0.0000	0.0000	0.0003	0.0104	0.0044	0.3931	0.3654	0.6269
SPANE_Neg	0.9111	0.3427	0.1013	0.0009		0.0018	0.0568	0.0000	0.0129	0.0000	0.2529	0.9721	0.0076
SWLS	0.7391	0.5169	0.0135	0.0000	0.0018		0.0000	0.0009	0.0005	0.0004	0.2082	0.0024	0.6266
Flourishing	0.5389	0.3764	0.0008	0.0000	0.0568	0.0000		0.9778	0.0000	0.6385	0.1025	0.9306	0.1379
Anxiety	0.8621	0.9511	0.4606	0.0003	0.0000	0.0009	0.9778		0.0279	0.0000	0.3905	0.6120	0.0033
SocialConnection	0.5090	0.7842	0.0009	0.0104	0.0129	0.0005	0.0000	0.0279		0.0069	0.0038	0.2553	0.0075
CESD	0.7217	0.9349	0.1047	0.0044	0.0000	0.0004	0.6385	0.0000	0.0069		0.0048	0.3840	0.0010
BelongSFU	0.2972	0.4537	0.0001	0.3931	0.2529	0.2082	0.1025	0.3905	0.0038	0.0048		0.2435	0.2176
eGPA	0.0559	0.8694	0.9116	0.3654	0.9721	0.0024	0.9306	0.6120	0.2553	0.3840	0.2435		0.0067
FEMALE	0.1754	0.7633	0.4382	0.6269	0.0076	0.6266	0.1379	0.0033	0.0075	0.0010	0.2176	0.0067	

Figure A1. The top matrix provides the raw correlation coefficient between measurements. The bottom matrix provides the p-value of a t-test of the null hypothesis that the correlation is zero. Stars are omitted, p-values closer to zero suggest statistically stronger correlations, e.g., the correlation coefficient between FEMALE and eGPA is positive ($\rho = 0.32$) and our test provides a p-value of 0.0067, replicating many findings that girls have higher high school GPAs. Note: the data in these tables is restricted to sample of individuals measured twice with complete responses to avoid the different samples affecting the correlation coefficient on POST.