Essays in College Course-Taking

by

Julian Hsu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2017

Doctoral Committee:

Professor Jeffrey Andrew Smith, Co-Chair
Assistant Professor Kevin M. Stange, Co-Chair
Professor John Bound
Professor Brian P. McCall

Julian Hsu
hjulian@umich.edu
ORCID ID: 0000-0002-6598-3592

DEDICATION

To my father, who continues to teach me the neverending joys of erring and learning, to double-check my work, to build (and avoid!) redundancies, and to live each day without regret.

To Carrie Wenjing Xu. Her endless love and support teaches me again and again that research is a human (and networked) endeavor.

Contents

List of Tables

List of Figures

List of Appendices

ABSTRACT

While college major choices are consequential, students do not know their capacity to perform in academic and non-academic environments related to those majors, and may not be prepared to take the classes for those majors. In my dissertation, I use detailed course transcript data to understand how students explore different majors and the roles of current and future college policies.

In my first chapter, *Learning about College Major Match: Microfoundations from Dynamic Course-Taking*, I develop and estimate an economic model to understand how first year courses can set students onto the path to their major. I develop and estimate a structural dynamic course-taking model that highlights how students learn about major match quality and complete course requirements to graduate in different majors. The model highlights how the breadth and depth of coursetaking across and within majors affects graduating major and graduation time. Estimating my model, I simulate a policy requiring students to take a variety of courses across different majors during their first year. As a more rigorous version of current colleges' distributional course requirements, this counterfactual policy causes the share of Natural Science graduates to increase ten percentage points. I find this counterfactual causes additional path dependence in the Natural Sciences from completing course requirements, rather than providing additional information.

My second chapter, *Do Grades Matter? Evidence from College Transcripts*, complements the first and dives into the correspondence between courses and majors. I combine administrative transcript data from a large public four-year institution to create a novel measurement of how student's progress in majors' course requirements. I find that four semesters after declaring one major, all students complete between 15% to 25% of the course requirements necessary to graduate in the Humanities, Social Sciences, and Psychology. Students' course-taking also seem to respond to first year grades. I construct a dynamic course-taking model, where in students take courses to learn about major abilities and complete majors' course requirements. The transcript data is consistent with the model's result that major switching costs increase as they continue to complete the course requirements in one major.

The third chapter, *Math for All? Regression Discontinuity in Signals of Preparation for*

College Quantitative Coursework (with William J. Gehring), uses plausibly exogenous variation to evaluate how higher education institutions can influence student course-taking and major choice. College calculus courses can be a stumbling block in pursuit of some goals for under-prepared students. We study how student course-taking and major decisions at an elite public institution respond to recommendations to take Pre-Calculus or Calculus. Using a regression discontinuity framework to estimate Intent-to-Treat effects, we find that, among the least-prepared students, students with a tentative recommendation to take Pre-Calculus are 60% more likely to ever take Calculus than if they receive a definite recommendation to take Pre-Calculus. We find suggestive evidence these recommendations equalize course-taking and major completion outcomes in Economics, Statistics, Biology, and Chemistry. We do find, however, evidence that students with the least favorable recommendation are more likely to be diverted toward quantitative courses that do not count toward a major. Our work suggests inducing students to take Pre-Calculus or Calculus is insufficient to encourage them into quantitative majors.

Chapter 1

Learning about College Major Match: Microfoundations from Dynamic Course-Taking

1.1 Introduction

Large variation in earnings across college graduates' majors makes major choice one of the most consequential economic choices made by young adults (Altonji et al., 2015; Black et al., 2003; Hastings et al., 2013; Kirkebøen et al., 2016). Altonji et al. (2012) find that the descriptive difference in log earnings between those with bachelor's degrees in electrical engineering and education is comparable to the difference between college and high college graduates.[1] Despite this importance, indecision casts a large umbra: one-third of students start college undeclared (U.S. Department of Education, 2016).[2] Stinebrickner and Stinebrickner (2014b) find a substantial proportion of students make unexpected major changes, indicating learning about major match quality.

United States colleges are designed for students to learn about their capacity to perform in academic and non-academic environments, or major match quality.[3] Colleges provide students with hundreds of courses designed to inform major choice, and require students to take courses across majors with distributional course requirements. Yet there is policy room in distributional course requirements' flexibilities to choose among courses designed for non-majors and to satisfy them at any time during college. In this paper, I find that requiring students to take courses designed for majors during their first year affects major choice by decreasing the effort needed to graduate in majors rather than providing information about major match quality.

Previous works use major declarations to proxy for students' information sets and course-taking, confounding two mechanisms of major choice: major match quality and coursework.

[1]Among males in the 2009 American Community Survey, and conditional on observerable characteristics available in the american Community Survey.

[2]Students do not start college with an official major, or declare before entering, but are "undeclared" in that they do not have any majors in mind.

[3]Many other countries require students to simultaneously choose a college and major, prioritizing time-invariant match quality based on pre-college characteristics and the ability to use focused curricula.

This paper directly uses course-taking with a course-taking model to separate these two mechanisms. It focuses on how courses provide information and complete majors' course requirements. Completed course requirements create natural incentives to ignore new information about college major match (Arcidiacono et al., 2016; Israel, 2005), because the effort needed to finish is lower. I measure the effort needed to finish in majors with the amount of courses needed to complete that major's course requirements.

Focusing on the role of courses, this paper makes two contributions to the major choice literature. The first is a structural single-agent dynamic course-taking model where in students take courses to jointly maximize short-run payoffs from course-taking during college, and long-run major-specific graduation payoffs. Courses give students information about their major match quality and complete majors' course requirements. Students only graduate in majors they have completed course requirements in. The model captures the trade-off between how diversified course-taking leads to more informed choices, and concentrated course-taking requires less effort to graduate.

The second contribution is using administrative transcript data from a four-year public flagship institution to estimate the structural model and simulate counterfactual academic policies. I identify the courses needed to graduate in different majors. The model is identified from students taking courses to control their graduating major and graduation time. The model's learning parameters are identified from the correlation of student course-taking and earned grades.

The model is among the first to use the systematic correspondence between course-taking and major choice, making it suitable for testing course-taking policies. Policy interest in using already available courses to influence major choice makes the need for such analysis considerable. I simulate students taking courses across three major groups – Natural Sciences, Humanities, and Social Sciences – during their first year. This corresponds to a more stringent version of current distributional requirements, designed to inform students' major choice with counterfactual information about their major match qualities.

I find this counterfactual policy increases the share of Natural Sciences graduates from 16% to 26%. Both counterfactual information about major match quality and systematically decreasing the course-taking effort needed to complete majors' course requirements drive this result. To decompose this result, I simulate another counterfactual where students do not learn about their major match quality from their imposed first-year course-taking. Finding similar results on graduating majors, I conclude that first-year courses affect major choice through decreasing the course-taking effort needed to complete majors' course requirements rather than providing additional information.

Finding previously completed course requirements can be stronger than learned match quality is crucial to understanding the impacts of induced course-taking.[4] While courses designed for majors are likely more informative of major match quality, they give students an additional incentive to persist in a major. Students' first-year courses are relatively uninformative but influence major choice independent of match quality. Administrators wanting to use course-taking as a policy lever need to be wary of attributing impacts to counterfactual information rather than completing course requirements.

Section 2 highlights how the paper improves on the college major choice literature. Section 3 presents the institutional context and transcript data. Section 4 presents the structural dynamic course-taking model. Section 5 discusses the identification strategy and concerns. Section 6 presents the model estimates and counterfactual simulations of how exogenous first year course-taking affects major choice. Section 7 concludes with policy implications and areas for future research.

1.2 Literature Review

The literature on college major choice can be separated into identifying different components of major match quality: expected labor market earnings, pre-college experiences, and college experiences. Beffy et al. (2012) and Long et al. (2014) find variation in labor market earnings have limited impacts[5] on college major choice. Wiswall and Zafar (2013) and Baker et al. (2017) find earnings have a substantial role,[6] although, observables only account for 20% of the variation in their major choice. The limited role is not surprising, considering Arcidiacono et al. (2012b) find students tend to have more information about their own majors' earnings potential.

Mixed evidence on the role of earnings prompted additional work focused on separating pecuniary from non-pecuniary roles in major match quality. One portion of match quality comes before college. A large segment of the college major choice literature looks at how students choose their major based on pre-college characteristics, including gender (Alms et al., 2016; Brown and Corcoran, 1997; Griffith, 2010; Turner and Bowen, 1999), race (Darolia and Koedel, 2016), family (Anelli and Peri, 2015; Zafar, 2012), and high school peers (Anelli and Peri, 2016). Some of the match quality is related to college performance (Speer, 2017), and

[4]Princeton Review and US News and Report encourage students to take courses across different majors: https://www.princetonreview.com/college-advice/choosing-college-majors, https://www.usnews.com/education/blogs/twice-the-college-advice/2011/11/15/10-tips-for-choosing-the-right-major

[5]Beffy et al. (2012) find a 10 percent increase in earnings causes a 0.25 and 0.53 percentage point increase in Natural Sciences, and combined Humanities and Social Science majors using French data. Long et al. (2014) find larger estimates with a reduced form approach using historic variation in earnings from American Community Survey, with more narrowly defined majors.

[6]Wiswall and Zafar (2013) find larger elasticities, with a 1 percentage increase in beliefs of own earnings in a major increasing the probably of majoring by 1.6 percentage points. Baker et al. (2017) find that among community college students, a one percentage point increase in earnings increases the probability of graduating in different majors by 1.5 percentage points.

Avery et al. (2016) find that higher Advanced Placement scores positively effect major choice, consistent with students learning about match quality.

The main hurdle in this area is that before students graduate in a major, they take courses and may declare other majors. Arcidiacono (2004) uses a structural model identified from variation in grades relative to students' calculated match quality beliefs, and Stinebrickner and Stinebrickner (2014b) uses students' surveyed expectations for declaring and finishing in different majors. Both find that students' beliefs about their major match qualities, independent from earnings, affect their major choice. I follow this literature using a structural model and allow correlated major match qualities.

Arcidiacono (2004) implicitly assume students focus their course-taking into a declared major, while Wiswall and Zafar (2013) assume course-taking is conditionally independent of major choice. Yet students take diverse course-taking trajectories through majors (Goldhaber et al., 2015; Hsu, 2017).[7] Then all but the last major declarations are noisy measures of course-taking, which are observable in transcript data and better measure how students learn and face effort costs during college[8].

Previous works modeling major declarations only used students' last declared majors to measure the effort needed to graduate in a major. This paper uses course-taking to model the effort[9] needed to graduate in different majors. Modeling courses makes students' entire course-taking history relevant and introduces additional dynamics for how students change their course-taking. Stinebrickner and Stinebrickner (2014b) use students' initial major interest and final major choice, note modeling course choices can become computationally intensive in tracking different combinations of choices, and requires additional assumptions. This is the first paper to tackle these issues head-on, incorporating how students learn from different courses[10] and the implicit costs of taking courses to learn about major match quality. While course-taking is the primary way to learn about major match quality, it also affects switching costs. This creates an unavoidable friction: taking courses in the Natural Sciences provides information and increases the cost of switching to other majors because they require relatively more courses to graduate. Switching costs then act as an information friction (Altonji et al., 2015); as the student gets closer to finishing a major's course requirements, the incentive to graduate looms larger.

Finally, this paper complements the literature on higher education design and whether stu-

[7](Andrews et al., 2014) finds similar patterns with students transferring between two- and four-year colleges.

[8]There is substantial work on the role of supply-side factors, such as instructors (Griffith, 2014; Price, 2010) and grade distributions (Bar et al., 2009).

[9]Baker et al. (2017) find that students' expected difficulty of future coursework also affects community college students' majors.

[10]Where Wiswall and Zafar (2013) and Stinebrickner and Stinebrickner (2014b) elicit students' expectations about their future major declarations across different majors, the dynamic course-taking model calculates students' expectations of future course-taking based on their course and grade histories. Relaxing this assumption greatly increases students' choices from declaring one major at a time to choosing courses across multiple majors.

dents should choose a major before college. Bordon and Fu (2015) and Malamud (2010, 2011) find students benefit from exploring majors before declaring during college, relative to choosing a major before college. Courses likely provide the most amount of information about major match quality, and modeling course-taking provides a microfoundation for how students explore majors.

1.3 Institutional Context: the University and Transcript Data

I use administrative transcript data for all students who enter a four-year public flagship institution, anonymized as the University, through its College of Arts and Liberal Sciences (CALS) from 2002 to 2011.[11] I combine transcript data with institutional information to measure how students complete different majors' course requirements.[12] This cannot be done with nationally representative data such as the National Longitudinal Survey of Youth (NLSY) or Beginning Postsecondary Students Longitudinal Study (BPS) for two reasons: insufficient sample size within institutions by major, and courses are reclassified with the Classification of Instructional Programs, making it difficult to identify individual courses that satisfy majors' course requirements.

This paper focuses on how students choose a major within the University's College of Arts and Liberal Sciences (CALS). CALS has the largest (60%) enrollment and has the most diverse set of majors.[13] CALS students interested in graduating in other majors must use an internal transfer process. Besides CALS, enrolling freshmen at the University also enter the Engineering, Medical-Related, or Arts-Related schools.

The data contains student administrative data on students' previous college exam scores and demographic data. To create equivalent math and reading standardized score measures, I use the percentiles in the SAT and ACT Math and Reading sections, averaging if both are available.[14] Also included are students' interests in different majors before starting any coursework at the University, reported household income, and reported maximum parental educational attainment.[15] Students can list interests in multiple majors, where in most nationally representative

[11] The transcript data follows students up until 2015. Although this departs from the standard convention of giving students six years to graduate, Table A.1 shows that more than 80% of students only attend the University for four years.

[12] Hendricks and Leukina (2015) use Postsecondary Education Transcript Study data to study student progress to completing general graduation requirements, rather than individual majors'. Bettinger (2010) uses transcript data without grades to examine trends in science course-taking. I discuss identification issues from ignoring the role of courses that do not satisfy any majors' course requirements in Section 1.5.3, and how requirements are coded in Appendix Section A.1.8.

[13] The University provides a contrasting environment to Berea College (Stinebrickner and Stinebrickner, 2014a,b), which focuses on helping underprivileged and talented students, and does not charge tuition, and to NYU (Wiswall and Zafar, 2013) and Duke University (Arcidiacono et al., 2014), which are private institutions.

[14] In the sample, around 20% of students only have SAT Scores, 55% only have ACT scores, and the remaining (except for 33, who are dropped in sample selection) have both.

[15] Previous academic interests come from the Common Application, and registering for the ACT and SAT. These questionnaires and discussion of how these interests are used can be found in Section A.1.2. Reported parental

datasets students can only list one.

One data drawback is the lack of post-graduation outcomes, such as earnings or employment. Without these data, I rely on estimating students' combined pecuniary and non-pecuniary payoff from graduating in different majors. I discuss conceptual difficulties in using earnings data from external datasets such as the ACS or NLSY97 in Section A.1.3.

1.3.1 Coding Majors' Course Requirements and Aggregating Majors

To track students' progress towards completing different majors' course requirements, I code CALS' majors' course requirements into the transcript data. Coding these requirements, I observe how much progress students make towards finishing each major's course requirements. Course-taking in CALS is not limited by declared majors, so students do not have an incentive to declare a major when they want to take courses in it.[16]

I standardize how students progress in majors' course requirements[17] into a percentage, where completing all the requirements is 100% progress, at the cost of being unable to identify individual courses.

CALS offers more than fifty different majors, and several have overlapping course requirements. To simplify the analysis, I create three different major groups: Natural Sciences, Humanities, and Social Sciences.[18] There does not exist a course that completes requirements in more than one major group.

Table 1.1 lists the individual majors within each major group. The three major groups are not as extensive as those used in previous works, because CALS simply does not offer them.[19] I include other majors commonly used in these works, such as Business, Engineering, and Education as internal transfer options for CALS students. At the University, the student cannot freely internally transfer into different colleges. The Business College requires a set of

characteristics come from the Common Application.

[16]Courses only satisfy major requirements if the student receives at least a "C." I observe most grades are above "C" : 3.67% of all courses taken earn at most a "C," and a separate 1.5% of all taken courses result in withdraws. Section A.1.4 provides a more thorough discussion of how major requirements were coded.

[17]The University offers academic minors as well, typically requiring eight to nine courses. While these minors are listed on students' graduation diplomas, they do not substitute for any academic majors and do not contribute towards graduation. The course requirements for minors are often a subset of those required for majors. I do not incorporate academic minors into this paper, and implicitly assume that there is no additional payoff from completing minors' course requirements.

[18]To create the aggregate progress towards completion in each major group, I use the maximum of progress of all the individual majors within that major group. Other statistics such as the average or median would make it difficult to measure a student completing 100% in that major. Another possible measure is to use the progress from the "relevant major." Say that a student has already made 10% progress in Biology and 20% in Chemistry within the Natural Sciences major group. If the student makes 10% progress in Biology and 5% progress in Chemistry in the next year, then I would record the 5% as the "relevant major" as it follows the major the student is closest to finishing. However, this under measures the students' effort in the Natural Sciences major group.

[19]Arcidiacono (2004) includes Business and Education. Wiswall and Zafar (2013) includes Business and Engineering. Stinebrickner and Stinebrickner (2014a,b) includes four more majors: Agriculture, Business, Elementary Education, and Professional Programs.

Economics courses, the Engineering College requires several Math, Physics, and other science courses, and other colleges have similar course requirements.

1.3.2 Trends in Completing Major Groups' Course Requirements

Even though all CALS students face the same general graduation course requirements, they engage major groups differently. Figure 1.1 finds variation in how different graduates complete major groups' course requirements over twelve semesters. Natural Science graduates make more progress in other major groups than other graduates, while Humanities graduates make the least. This is consistent with Natural Science graduates exploring other majors to a greater depth than Humanities graduates.

Regardless of graduating major, the average graduate has taken a substantial amount of courses across the different major groups. Hsu (2017) shows there is a loose correspondence between such course-taking and major declarations. Focusing on major declaration would under-measure how much information and experience students have across major groups.

There is very little variation after students' eighth semesters, and I limit analysis to students first four years of enrollment. To understand how students are taking courses to satisfy different major groups' course requirements, I create a dynamic course-taking model.

1.4 Dynamic Course-Taking Model

The dynamic course-taking model relaxes a core assumption used in the college major choice literature: instead of declaring one major or another, the student allocates discrete combinations of requirement units towards completing different major groups' course requirements. Requirement units abstractly represent her course-taking and how she completes major groups' course requirements. If she allocates at least four requirement units into a major group, she graduates in that major group.[20] She trades off diversifying and concentrating her requirement unit allocations to jointly maximize her payoffs from course-taking in college and graduation payoffs.

The remainder of the model follows those in the college major choice literature: the student uses grades she receives upon allocating requirement units into different major groups to learn about her major group match qualities. The student also believes her major group match qualities are correlated.

[20]Tracking the student's completed requirements across majors also incorporates path dependence into the student's choice. If she is closer to finishing a Natural Sciences major group she has more incentive to finish it. Allowing for a more flexible choice set, the model is similar to a multi-armed bandit problem (Gittins, 1979; Whittle, 1988): the student can allocate requirement units across multiple majors, learn about her major group match qualities, and continues allocating until graduation.

1.4.1 Choice Structure: Requirement Unit Allocations, Internal Transfers, and Drop Out

In each academic year t in T, the student i considers different combinations of requirement units to allocate across M major groups. Each requirement unit completes 25% of a major group's course requirements.

The student can allocate at most three requirement units in any of the combinations shown below. She cannot allocate all three requirement units into (complete 75% of the course requirements for) one major group in one time period. The model treats this as exogenous: in reality it is likely due to scheduling conflicts (three requirement units corresponds to eight to ten courses in one year).

All Possible Requirement Unit Allocations, Set $\mathbb{L}$:

$(e_{\text{i, Natural Sciences, t}}, e_{\text{i, Humanities, t}}, e_{\text{i, Social Sciences, t}}) =$

(1,0,0)	(2,1,0)	(1,0,1)	(0,0,2)
(2,0,0)	(0,2,0)	(2,0,1)	(1,0,2)
(0,1,0)	(1,2,0)	(0,1,1)	(0,1,2)
(1,1,0)	(0,0,1)	(0,2,1)	(1,1,1)

When the student chooses one combination of requirement unit allocations, $\mathbb{c} \in \mathbb{L}$, she receives an immediate course-taking payoff $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$. If she accumulates at least four requirement units in major m, she graduates in major group m the next period and receives graduation payoffs W_{imt}. If she simultaneously finishes multiple major groups' course requirements in one time period, she also receives a combination-specific graduation payoff.

The student can also internally transfer into separate Business College or Other College. Internally transferring to other colleges, graduating in a major group, and dropping out are the only terminal states. Her entire choice set $\mathbb{C}$ is internally transferring to these two colleges, dropping out, and combinations of requirement units $\mathbb{L}$. Her objective can be briefly described with the following Bellman equation:

$$
\begin{aligned}
V(S_{it}) \quad = \max \Big\{ & \max_{\mathbb{c} \in \mathbb{L}} \Big\{ \nu(\mathbb{c}|\mathbb{c} \in \mathbb{L}) + \mathbb{1}\{\text{Graduate in None}|\mathbb{c}, S_{it}\}\beta\mathbb{E}[V(S_{it+1}|\mathbb{c})] + \\
& \mathbb{1}\{\text{Graduate in } m|\mathbb{c}, S_{it}\}\beta W_{imt} \Big\}, \\
& u^{Bus.}, \\
& u^{Oth.}, \\
& \text{Drop Out}, \Big\}
\end{aligned} \tag{1.1}
$$

The student's state variables – completed requirements and grade point averages (GPA) across major groups – are S_{it}. The state variables are used to calculate the student's expectation of future payoffs and match quality beliefs. She considers the probability of different possible future grades and idiosyncratic shocks to future payoffs to calculate her expected future immediate college flow payoffs from each choice $\mathbb{c}$ and future graduation payoffs.

1.4.2 Immediate Course-Taking Payoffs

I parameterize $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$ using her requirement unit allocation: $\mathbb{c}_{it} = (\mathbb{c}_{i \text{ NatSci } t}, \mathbb{c}_{i \text{ Human } t}, \mathbb{c}_{i \text{ SocSci } t})$.

The immediate payoff $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$ is linearly separable into major-group-specific payoffs, u_{imt}, which depend on how many requirement units she allocates within one major group, her major group match quality, and time-invariant individual characteristics Z_i. The immediate payoff also has an interaction payoff $u(\mathbb{c})$ that depends on how she allocates requirement units across major groups.

$$\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L}) = \sum_{m \in \mathbb{c}} u_{imt}(b_{imt}, \mathbb{c}_{imt}, \text{Pre-College Interest in } m_i, Z_i) + u(\mathbb{c}_{it}) \tag{1.2}$$

The student receives u_{imt} from any major group m she allocates at least one requirement unit into. The student does not know her major group match qualities, and relies on her belief in them, b_{imt}. Match quality appears in how the student benefits from discussing the material with peers, expected grades, and effort costs. Payoffs from match quality vary over major groups. For example, having high match quality may be more valuable in the Natural Sciences than Humanities.

Parameterizing u_{imt} captures two ways payoffs vary over course-taking. First, the student's effort can increase concavely or convexly as she takes more courses in major group m. I use fixed effects for allocating one or two requirement units to capture non-linearities in effort. Second, elective courses are differently designed from introductory courses. The student receives an additional payoff after allocating two requirement units in that major group – an approximation for taking advanced elective courses. Third, the student's course-taking experience simply scales with her course-taking. Her payoff with respect to her match quality beliefs b_{imt} and individual characteristics Z_i linearly scale with her allocation $\mathbb{c}_{imt}$. Z_i proxy for the student's familiarity with the major group, and I use indicators of interest in the major group before college, and indicators of US citizenship, reported household income, and maximum parental education.

$$\begin{aligned} u_{imt} = \quad & \kappa_{1m} \mathbb{1}\{\mathbb{c}_{imt} = 1\} + \kappa_{2m} \mathbb{1}\{\mathbb{c}_{imt} = 2\} + \\ & \kappa_{\text{adv},m} \, \text{freqadv}(\mathbb{c}_{imt}) + \mathbb{c}_{imt} \Big(\kappa_{b,m} b_{imt} + \kappa_{z,m} Z_i \Big) \end{aligned} \tag{1.3}$$

where $\text{freqadv}(\mathbb{c}_{imt})$ counts the number of requirement unit allocations the student makes after the first two requirement units.

Aside from how the student completes the requirements within a major group, it is also important how the student allocates requirement units across major groups. The interaction payoff

$u(\mathbb{c})$ captures how the student's payoff changes non-linearly if she concentrates or diversifies her requirement unit allocations within one time period. The student's effort likely increases if she takes courses across multiple major groups, and I have indicators of whether the student is taking requirement units in one or more major groups. Besides being able to allocate requirement units across major groups, the student can also allocate more than one requirement unit into a major group. I use the variance of $\mathbb{c}$ to account for how this second margin affects payoffs.

$$
\begin{aligned}
u(\mathbb{c}) \quad &= \pi_1 \mathbb{1}\{\tilde{\mathbb{c}_{it}} = 1\} + \pi_2 \mathbb{1}\{\tilde{\mathbb{c}_{it}} > 1\} + \pi_{var}\,\text{variance}(\mathbb{c}) + \epsilon_{i\mathbb{c}t} \\
&\text{where } \tilde{\mathbb{c}_{it}} = \sum_{\tilde{m}=1}^{M} \mathbb{1}\{\mathbb{c}_{i\tilde{m}t} > 0\}
\end{aligned}
\tag{1.4}
$$

Aside from how the student allocates requirement units across major groups, unobserved payoffs to course-taking such as inconvenient scheduling, taking courses with friends, and availability of certain courses can idiosyncratically affect the immediate payoff. $\epsilon_{i\mathbb{c}t}$ captures these idiosyncratic shocks. It is likely components of $\epsilon_{i\mathbb{c}t}$ are common across major groups, such as major group specific shocks that affect some requirement unit allocations and not others. $\epsilon_{i\mathbb{c}t}$ is distributed over a generalized extreme value distribution, with location and scale parameters 0 and τ_t and correlated across requirement unit allocations $\mathbb{L}$ with correlation $\varphi_{\mathbb{L}}$.

1.4.3 Internal Transfer and Graduation Payoffs

Beyond allocating requirement units towards completing major groups' course requirements in CALS, the student can also internally transfer to a Business College or Other College or drop out. The drop out payoff is normalized to zero. If the student internally transfers to the Business College or Other College within the University, she receives payoffs of $u^{Bus.}$ or $u^{Oth.}$.

These payoffs vary over when the student starts at the Business College or Other College, since all students are required to fulfill the same specialized course curricula regardless of entering year. Then fixed effects of when the student enters the Business College or Other College during her second, third, or fourth year represents the combined effects of starting the specialized curricula and potentially graduating later if she enters in her third or fourth year.

I allow $u^{Bus.}$ and $u^{Oth.}$ to vary over students' GPAs in the three major groups, and are reduced form representations for the expected value of successfully applying, internally transferring, and (most likely) graduating in the Business College or Other College.[21] The payoff also varies over whether the student expressed interest in these Colleges before enrolling at the

[21] Once she leaves CALS, I assume the student no longer learns about major group match qualities. An interesting avenue for future work is how students decide more specialized fields within each of these colleges.

University, which proxies for individual motivation independent of GPAs.

$$
\begin{aligned}
u^{Bus.} \quad & = \kappa^{\text{Bus.}}_{t=2} + \kappa^{\text{Bus.}}_{t=3} + \kappa^{\text{Bus.}}_{t=4} + \\
& \kappa^{\text{Bus.}}_{1}\overline{g_{i,\text{ NatSci, }t}} + \kappa^{\text{Bus.}}_{2}\overline{g_{i,\text{ Human, }t}} + \kappa^{\text{Bus.}}_{3}\overline{g_{i,\text{ SocSci, }t}} + \\
& \kappa^{\text{Bus.}}_{\text{int}}\text{Pre-College Interest in Bus.}_{i} + \epsilon_{i\mathbb{C}t} \\
u^{Oth.} \quad & = \kappa^{\text{Oth.}}_{t=2} + \kappa^{\text{Oth.}}_{t=3} + \kappa^{\text{Oth.}}_{t=4} + \\
& \kappa^{\text{Oth.}}_{1}\overline{g_{i,\text{ NatSci, }t}} + \kappa^{\text{Oth.}}_{2}\overline{g_{i,\text{ Human, }t}} + \kappa^{\text{Oth.}}_{3}\overline{g_{i,\text{ SocSci, }t}} + \\
& \kappa^{\text{Oth.}}_{\text{int}}\text{Pre-College Interest in Oth.}_{i} + \epsilon_{i\mathbb{C}t}
\end{aligned}
\tag{1.5}
$$

$u^{Oth.}$ and $u^{Oth.}$ also have idiosyncratic shocks $\epsilon_{i\mathbb{C}t}$ that are distributed extreme value type 1 with location and scale parameters 0 and τ_t. These idiosyncratic shocks are not correlated with those in the immediate payoffs, and represent unobserved shifts in the application process or student interest in transferring to the Business College or Other Colleges.

The student graduates from a major group in CALS if she accumulates at least four requirement units into any major group. I denote graduating in major group m as K_m.[22] She receives W_{imt}, which represents her expected future discounted sum of pecuniary and non-pecuniary payoffs after graduating in major group m. W_{imt} depends on her graduating major-group GPA, $\overline{g_{imt}}$. Since she graduates after realizing her final grades, she relies on her expectation of her final GPA in major group m.

I also allow graduation payoffs to vary over entering cohort years to account for graduation trends. Graduation trends can come from varying labor market returns to different majors, such as during the Great Recession from 2007 to 2009.[23]

$$
\begin{aligned}
W_{imt} = \quad & \rho_{m,t=3} + \rho_{m,t=4} + \\
& \rho_{g,m}\overline{g_{imt}} + \rho_{1m}\text{2006 to 2007 Cohort}_i + \rho_{2m}\text{2008 to 2011 Cohort}_i
\end{aligned}
\tag{1.6}
$$

$\rho_{m,t=3}$ and $\rho_{m,t=4}$ represent the combined effect of graduation time on tuition and future earnings. Graduating earlier means paying fewer years of tuition and having additional years to accrue graduation payoffs. Graduating earlier likely gives employers a positive signal as well. The University can also influence students' graduation times through policies such as requiring students to declare a major after their sophomore year. The time fixed effects are a reduced form representation of these different mechanisms. Section 1.5.3 discusses possible misspecification issues.

Finally, if the student finishes multiple major groups' course requirements in the same time period, she receives the average of these major groups, as well as an additional combination-specific payoff. Hemelt (2010) and Rossi and Hersch (2016) find limited labor market returns

[22]Formally, K_m is the condition where $\sum_{t=1}^{T'} \mathbb{c}_{imt} \geq 4$.

[23]Timespan comes from the National Bureau of Economics Research: http://www.nber.org/cycles.html

to double-majoring.

$$\begin{aligned}
\widetilde{W}_{i \text{ NatSci \& Human } t} &= \tfrac{1}{2}W_{i \text{ NatSci } t} + \tfrac{1}{2}W_{i \text{ Human } t} + \rho_{\text{NatSci \& Human}} \\
\widetilde{W}_{i \text{ NatSci \& SocSci } t} &= \tfrac{1}{2}W_{i \text{ NatSci } t} + \tfrac{1}{2}W_{i \text{ SocSci } t} + \rho_{\text{NatSci \& SocSci}} \\
\widetilde{W}_{i \text{ Human \& SocSci } t} &= \tfrac{1}{2}W_{i \text{ Human } t} + \tfrac{1}{2}W_{i \text{ SocSci } t} + \rho_{\text{Human \& SocSci}} \\
\widetilde{W}_{i \text{ NatSci \& Human \& SocSci } t} &= \tfrac{1}{3}W_{i \text{ NatSci } t} + \tfrac{1}{3}W_{i \text{ Human } t} + \\
&\quad \tfrac{1}{3}W_{i \text{ SocSci } t} + \rho_{\text{NatSci \& Human \& SocSci}}
\end{aligned} \tag{1.7}$$

The student's major group match qualities indirectly affect her graduation payoff through her expectation of graduating GPA. Her beliefs, b_{imt}, directly impact her immediate payoffs from allocating requirement units. Learning about match qualities from earned grades link the uncertainty in future grades with learned major group match quality. I describe the learning framework below.

1.4.4 Learning About Major Group Match Qualities

Before the first period, the student is endowed with immutable major group match qualities $\mu_i = (\mu_{i \text{ NatSci}}, \mu_{i \text{ Human}}, \mu_{i \text{ SocSci}})$. μ_i are drawn from a multivariable normal distribution, with mean Γ and covariance matrix Δ. She does not know the values of μ_i, and has a prior beliefs, b_{i1}, about them based on pre-college characteristics, X_i. X_i includes students' ACT and SAT percentile scores, reported gender, and reported race. As she receives grades, g_{it}, and allocates requirement units, she Bayesian updates her major group match quality beliefs $b_{it} = (b_{i \text{ NatSci } t}, b_{i \text{ Human } t}, b_{i \text{ SocSci } t})$. I assume the student has rational expectations over future grade realizations and corresponding belief changes.[24]

$$b_{im1} = \phi_m + \phi_m X_i + \epsilon^b_{im} \tag{1.8}$$

$$\begin{aligned}
g_{imt} &= \mu_{im} + \eta_{imt} \\
\eta_{imt} &\sim \mathrm{N}(0, \sigma^2_m)
\end{aligned} \tag{1.9}$$

She uses her prior beliefs b_{i1}, cumulative progress and GPA to Bayesian update her beliefs after receiving immediate course-taking payoffs $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$ and before making her next choice.

The student believes the expectation error between her expected grade, $\mathbb{E}[g_{imt}] = b_{imt}$, and actual grade, g_{imt}, comes from not knowing μ_{im} and idiosyncratic noise η_{imt}. In the Bayesian updating process, the student weights the grades she receives with her previously allocated requirement units. As she earns more grades and learns more about her match quality, she attributes more of the expectation error to be idiosyncratic noise.

[24]The rational expectation and Bayesian learning assumptions are not innocuous. Rational expectations implies that across the population, students' expectations of their match qualities align with the population average – on average, students' beliefs are correct. Stinebrickner and Stinebrickner (2014b) find differences in students' beliefs from the population average, and use students' elicited beliefs. Without observing beliefs in the data, I rely on these assumptions to calculate students' beliefs.

Since the student's abilities μ_i are distributed multivariate normal with covariance Δ, the student knows match qualities μ_i are correlated, and uses Δ in the learning process. With this correlated framework, the student updates her beliefs in major groups $k \neq m$ from a grade in major group m. She forms her beliefs as follows (DeGroot, 1979):

$$\begin{aligned} \mathbb{E}_t[\mu_i] = b_{it} &= \mathbb{V}_t[\mu_{it}] \cdot \Big((\Delta)^{-1} b_{i1} + (\Sigma)^{-1} \sum_{t'=1}^{t} g_{it'}\Big) \\ \mathbb{V}_t[\mu_i] &= \Big((\Delta)^{-1} + (\Sigma)^{-1}(\sum_{t'=1}^{t} \mathbb{c}_{it})\Big)^{-1} \end{aligned} \tag{1.10}$$

where the diagonal matrix Σ has diagonal entries σ_m^2.

There are information gains to allocating requirement units across major groups. The choice to learn about her match qualities can come at the cost of delaying her graduation – she trades off short-run gains from graduating earlier with potentially higher payoffs from learning about match qualities. The student can exploit the correlation between major groups to learn more quickly, without allocating requirement units into them.

1.4.5 Implications for Information-Seeking Behavior and Switching Costs

This learning framework builds concavity into the student's requirement unit allocation choice, because there is an informational risk to allocating two requirement units into a major group. The student cannot receive multiple grades from the major group in one period,[25] and weights the grade with the total requirement units she allocated in that major group. She risks a large change to her ability beliefs[26] if she allocates two requirement units and receives a low grade. One way to avoid this is to diversify her requirement units across major groups to learn more slowly.

Yet for certain students, this risk is low if match quality beliefs are sufficiently high. Students with higher match quality beliefs should allocate more requirement units to exploit higher expected payoffs. Students with lower match quality beliefs have an incentive not to learn about, allocate requirement units in, and graduate in those major groups. Goldhaber et al. (2015) find students with higher SAT scores focus more of their course-taking.

A final dynamic is independent of learned match quality and can act against learned match quality. A student who has concentrated her requirement units in a major group has a greater incentive to continue because she is closer to graduating. The implicit switching cost of gradu-

[25] Related to the assumption all courses are identical, keeping track of the course or requirement unit weighted grade the student receives within a time period creates additional computational burden. Section A.4.3 elaborates on this.

[26] Even if the student perfectly knew her match qualities, she may not make the same choices each time period. Suppose she receives positive immediate payoffs from allocating requirement units to Natural Sciences, but does not intend to graduate in it. She actually wants to graduate in Social Sciences because it has a higher graduation payoff. Then one way for her to maximize her utility is to allocate requirement units in Natural Sciences and Social Sciences, and in the last period, only allocate enough to graduate in the Social Sciences.

ating in one major group grows as the student allocates more requirement units into other major groups. The student stands to potentially graduate later and incur effort costs from allocating additional requirement units. Switching costs become crucial if the student learns she has low match quality in a major group, but only needs one more requirement unit to graduate. She has an incentive to ignore learned match quality to graduate earlier.

1.5 Estimation Strategy and Identification

I estimate model parameters to match the likelihood students make each of the observed choices and receive observed grades using maximum likelihood estimation. Model parameters are identified through the share of students who make different choices over time. Cross-sectional variation in student's characteristics and panel variation in cumulative requirement units and GPA across major groups identify major-group-specific payoff parameters in u_{imt}. Graduation payoff parameters are identified from how students consistently allocate requirement units in a major group over time, and the share of students who graduate over time. Learning parameters are identified from the differences between realized grades and calculated match quality beliefs.

In this section, I also discuss identification issues that arise from assuming only courses that satisfy major groups' course requirements give students information and assuming that the University's other course requirements do not affect students' decisions.

In order to calculate the probability students make different choices, I use backwards induction to solve the Bellman equation in (1.1). To address computational burden in calculating the value functions across different combinations of requirement units and GPAs, I estimate the model using a random subset of 4,000 students from the final sample.

1.5.1 Forming the Choice Set

Within each year, I discretize how students complete major groups' course requirements into bins of 25%.[27] Each period is one year: I take advantage of the steep drop in attending students after their fourth year as shown in Table A.1 and aggregate all course-taking from their fourth year onwards into one year. This is to use a tractable choice set for estimation. A key feature of the dynamic course-taking model is tracking how students complete major groups' course requirements over time. As I increase the number of time periods or number of requirement unit allocations, the number of possible state variable values exponentially increases.

Discretizing completing course requirements can over-estimate the share of students who are graduating in each major group. For example, I could infer that a student who takes one course in the Natural Sciences in each year graduates in the Natural Sciences. To avoid this, I set 10% as the minimum amount of progress to count as making 25% progress in a major

[27] Figure A.1 compares the distributions of how students actually complete major groups' course requirements in each academic year, with the discretized requirement units. The discretization adequately represents how few students complete many course requirements in an academic year.

group.

Table A.10 shows that more than 99% of observed student-year requirement unit allocations share two features:

1. Allocate between one and three requirement units across all major groups
2. At most allocate two requirement units in one major group.

It is likely few students complete 75% of the course requirements in one major group (corresponding to eighth to ten courses in a year) due to scheduling conflicts. Using this discretized measure, I track how students complete major groups' course requirements over four years.[28]

1.5.2 Sample Selection

I focus on a sample of students with similar prior college experiences who make the observed choices in the course-taking model. Table 1.2 shows that I first drop around 12% (4851 out of the beginning sample 41331) students due to empirical concerns – entering the University with more than 24 credits from other institutions or AP exams.

Students who transfer a substantial amount of credits from outside institutions start at CALS with more information about their major group match qualities and have already completed some major groups' course requirements. While the University does not accept all these credits towards different requirements, students with substantial transfer credits have academic experiences that are incomparable to those who start at the University.[29]

I then drop around 2% of the beginning sample (691 out of 41331) students because they make choices outside those allowed in the course-taking model. This includes students who allocate three requirement units in one major group and allocate more than three requirement units over all major groups in any given year. Again, accommodating these choices increases the computational burden of calculating the value functions across all combinations of previous requirement unit allocations and GPAs. I also place restrictions on whether students can internally transfer: students cannot internally transfer to the Business College or Other Col-

[28] In the model, students leave the University immediately after accumulating at least four requirement units in a major group. Therefore, I drop all student-year observations after the student is measured to graduate. I show how many student-year observations are dropped in Table A.7. Around 3,000 students are inferred to graduate from CALS before they actually do. Most of these students are coded as graduating in the Humanities or Social Sciences major groups in their third year but actually graduate in their fourth year. Besides these students, around 4,000 students from the final sample lose observations after they leave CALS because they internally transferred into Business College or Other Colleges. I consider internally transferring to the Business and Other Colleges as a terminal state, so this is not a concern for estimation.

[29] Around 7% of students dropped due to empirical concerns ever enroll at half-time. The decision to attend the University half-time instead of full-time is likely driven by external factors, since students taking a course load with fewer than 12 credits each semester face a different half-time tuition payment schedule and have considerably more time to spend on other priorities. I do not observe any financial aid or tuition expense in the data, and I drop students who ever enrolled half-time. Section A.2 discusses this in detail.

lege during their first year.[30] Section A.2 compares these dropped students' characteristics and outcomes in detail.

Table 1.2 shows CALS students stand out compared to most college students. 40% of students reported on their Common Applications that their households earn more than $100,000 per year, and 44% of students have at least one parent with an advanced (Masters, Doctorate, Medical, or Law) degree. Taking these proportions at face-value, these students come from more advantaged backgrounds than those at Berea College (Stinebrickner and Stinebrickner, 2014b), but have similar backgrounds to students at Duke University (Arcidiacono et al., 2012a) and New York University (Wiswall and Zafar, 2013).[31]

1.5.3 Parameter Identification and Bias Issues

I estimate the utility parameters in the immediate payoffs to allocating requirement units in (1.2), graduation payoffs in (1.6), and Bayesian learning parameters in (1.10). Utility parameters are identified relative to the drop out options normalized to zero.

The immediate course-taking payoffs $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$ are identified from the share of students who allocate different requirement units over time, and how students choose to stay in CALS to continue allocating requirement units in the future. The model infers that major groups with higher payoffs must be attracting more students, and students do not rush to graduate and stay longer in college because $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$ values are higher.

Estimates on how advanced requirement units in the immediate course-taking payoffs, $\kappa_{\text{adv},m}$, are identified from the share of students who continue to allocate requirement units in a major group after previously allocating two. $\kappa_{\text{adv},m}$ is an incentive for students to not only allocate two requirement units, but persist afterwards (this is separate from the identification for graduation payoffs).

I assume idiosyncratic shocks $\epsilon_{i\mathbb{c}t}$ are distributed generalized extreme value, with location parameter 0, scale parameter τ_t, and correlated $\varphi_{\mathbb{L}}$ between requirement units allocations.[32]

[30] I also drop the extremely small share of students who I measure as completing the course requirements for one of the major groups by their second year. There are 95 in the final sample of students. I use a random subset of the final sample, which exacerbates small share issues. It is also questionable whether students who transfer out of CALS in their first year had intentions of staying in CALS in the first place.

[31] Based on Arcidiacono et al. (2012a)'s Table 1, which is conditional on students who have Duke Admission Office evaluations, my calculations indicate 52% of students' households have annual earnings greater than $100,00. My calculations indicate 33% – 45% have at least one parents with an advanced degree. Wiswall and Zafar (2013)'s sample of New York University students' annual household earnings is $143,840, and 71% and 75% of mothers and fathers have a bachelors' degree, respectively.

[32] This correlation also addresses issues of having no major-group-specific idiosyncractic errors in the immediate payoffs. Major-group-specific errors would introduce mechanical correlations of $\epsilon_{i\mathbb{c}t}$ for combinations that allocate requirement units to the same major groups. Leaving out these major-specific errors introduces bias into estimating $\kappa_{b,m}$ since students' match quality beliefs vary over time. Introducing these major-group-specific idiosyncratic errors greatly increases the computational burden for using backwards induction to solve the value functions in (1.1), as it would require simulation-based methods to calculate the probability students make different choices.

The scale parameter is identified from the distribution of students' choice within an academic year, and their decision to stay in CALS, conditional on the other payoffs. To normalize utility levels across years, I normalize the scale parameter for $\epsilon_{i\mathbb{c}t}$ in the first period, $\tau_1 = 1$.

The continuation value[33] of staying in CALS, following McFadden (1978), increases in τ_t and $\varphi_{\mathbb{L}}$. τ_t determines the conditional distribution of requirement unit allocations. Intuitively, as variation in $\epsilon_{i\mathbb{c}t}$ increases, students' choices within a year are left to factors outside the model. $\varphi_{\mathbb{L}}$ is identified from the nested logit structure: the share of students who allocate requirement units rather than internally transfer or drop out.

One identification issue is ignoring the general course requirement to complete 120 credits to graduate at all. This 120 credit requirement gives students an additional incentive to take courses and stay in college longer. If the reason students are staying in CALS is to satisfy the 120 credit requirement,[34] then the estimates of the value of future choices $\mathbb{E}[V(S_{it+1})|S_{it}, \mathbb{c}]$, are upwards biased. This can operate through an upwardly biasing in estimates of immediate payoffs, $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$.[35]

Internal transfer payoffs are identified from cross-sectional variation in students of different GPAs who internally transfer to different Colleges. Individual variation comes from GPAs across major groups and pre-college interest in these Colleges. $u^{Bus.}$ and $u^{Oth.}$ are reduced form representations of the entire experience in the Business College and Other College, including the internal application process. It is likely unobserved factors about student motivation and Colleges' preferences for certain types of students positively correlate with GPA. Note that the idiosyncratic shocks $\epsilon_{i\mathbb{c}t}$ also appear in the internal transfer payoffs, which I assume are not correlated with $\varphi_{\mathbb{L}}$.

Graduation payoffs W_{imt} motivate students to allocate requirement units into a major group because it directly determines which graduation payoffs students receive. These payoffs discontinuously occur once the student has allocated four requirement units into a major group, while the immediate course-taking payoffs $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$ can be received each year the student is in CALS. The discontinuous change separately identifies utility parameters in graduation payoffs W_{imt} from immediate payoffs $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$.

Two sources of variation identify the graduation payoffs. The first is students who are "at risk" of graduating. Among students who have allocated three requirement units in a major group, the greater the share of students who allocate the fourth requirement unit to graduate,

[33]The analytical form for the expectation of the maximum of future choices is (A.2). Intuitively, it captures that a student may want to continue in CALS for the chance to receive a higher $\epsilon_{i\mathbb{c}t}$, even if all the parameters are negative.

[34]Consider a simple example. Suppose a student prefers to spend the first two years taking courses that is to satisfy the 120 credit requirement and the last two years taking courses that satisfy major groups' course requirements. This preference causes her to attend the University for four years instead of two years.

[35]Similarly, I assume that scheduling restrictions limit how much progress students can make in different major groups within a year. Table A.2 shows limited evidence that students treat courses that do and do not satisfy major groups' course requirements as substitutes.

the model infers a larger graduation payoff.

The second is how students allocate requirement units conditional on how they have already completed major groups' course requirements. If graduation payoffs are high, then students who have higher cumulative requirement units in a major group will allocate more requirement units in the next academic year. If students make different requirement unit allocations in each year, the model would infer graduation payoffs have little value.

It is possible the graduation payoffs are misspecified because I assume students' allocated requirement units and GPA in one major group do not affect another major group's graduation payoffs. This parameterization rules out a student taking courses in Computer Science (Natural Sciences) to increase her Economics (Social Sciences) graduation payoff. In this example, the model would infer the student receives a high immediate course-taking payoff from allocating requirement units into the Natural Sciences. Relaxing this assumption makes identification difficult: the model cannot differentiate between students allocating requirement units into a major group for the short-run immediate payoff or long-run graduation payoff.

W_{imt} is not identified using labor market earnings, and there is no idiosyncratic shock in W_{imt}. Individual variation comes from cohort fixed effects. Idiosyncratic shocks would represent unforeseen labor market changes or individual shocks to the expected value of graduating in different major groups. Arcidiacono et al. (2012b) suggests that students learn more about the earnings of their own majors. Then idiosyncratic shocks must be very large to play a substantial role, since W_{imt} represents the future discounted sum of payoffs after graduation.

A key component to the utility payoffs is students' beliefs about their major group match qualities. Although the internal transfer payoffs depend on current GPA, students planning on internally transferring in the next period have expectations of these GPAs. Match qualities directly enter the immediate course-taking payoff, and students use match quality beliefs to form expectations of graduating GPAs.

The identification assumption for the learning process in (1.10) is that the expectation error between grades and match quality belief is random.

$$\mathbb{E}[g_{imt} - b_{imt} | b_{imt}] \perp\!\!\!\perp b_{imt}$$

This expectation error only contains the error about beliefs and idiosyncratic grade errors η_{imt}. As students learn more about their match quality, they become more certain about them. They attribute more of the expectation error to η_{imt}, and are better able to predict their grades over time.

The off-diagonal elements covariance matrix Δ is identified from the correlation between major group GPAs and future choices. If students who allocate more requirement units in a major group and have higher GPAs in other major groups, then the model infers the correlation between major group match qualities is positive. The diagonal element of Δ are identified from the distribution of calculated match qualities.

There are two identification issues with the learning mechanisms. The first is that students could be learning about their match qualities from their course experiences, independent of their grades.[36] If non-grade match information is negatively correlated with grades, then estimates of $\kappa_{b,m}$ are downwards biased.

The second identification issue is only that students learn from courses that satisfy major groups' course requirements. The model explicitly assumes other courses provide no information. Consider a case where a student has taken Biology 201 which does not satisfy any major groups' course requirements. She has also taken Economics 101 which satisfies course requirements in the Social Sciences major group. If she allocates requirement units in the Natural Sciences because of her grade in Biology 201, the model would incorrectly infer a high correlation between Natural Sciences and Social Sciences.

1.6 Model Estimates and Counterfactual Simulations

1.6.1 Model Estimates

Model estimates on Table 1.4, Panel A, show the fixed effects for allocating one or two requirement units κ_{1m} and κ_{2m} are negative in the Natural Sciences, suggesting Natural Science courses are generally uncomfortable for students. The Humanities and Social Science major groups have positive estimates. I find $\kappa_{1m} > \kappa_{2m}$, consistent with convexly increasing effort costs. Estimates on the number of advanced requirement units – all requirement units allocated after the second – are positive. These advanced requirement units proxy for taking elective courses in different major groups, and the estimates suggest students enjoy these courses.

Relative to students who report less than \$100,000 annual household income, students who report more than \$100,000 annual household income receive lower payoffs in the Humanities and Natural Sciences. Students who did not report also receive lower payoffs. These estimates are neither statistically not substantially significant. Relative to students who report their maximum parental education is high school, students from more educated backgrounds receive higher payoffs from allocating requirement units in the Natural Sciences but lower payoffs in the Humanities and Social Sciences.

The estimated parameters that link match quality beliefs with immediate payoffs, $\kappa_{b,m}$, are positive only for the Natural Sciences: students with higher match quality beliefs in the Natural Sciences are more likely to allocate requirement units in the Natural Sciences. Interestingly, it is negative for the Humanities and Social Sciences. This differs from previous works, which

[36] Ackerberg (2003) and Crawford and Shum (2005) find evidence of taste learning in yogurt and pharmaceutical consumption, and consumers learn about their tastes for different goods.

largely find major declarations are positively related with GPAs and match quality beliefs.[37] Table 1.3 shows there is more variation in whether students allocate one or two requirement units in the Humanities and Social Sciences, than whether they allocate any.[38] This suggests these estimates are identified off the intensive rather than the extensive margins, and I interpret the negative estimates for Humanities and Social Sciences as students being more likely to allocate one rather than two requirement units.

Estimates of graduation payoffs in Panel B are consistent with previous findings of positive correlations between GPAs and labor market outcomes. The estimates from graduating in multiple majors are positive, consistent with students wanting to graduate in different major combinations. However, the estimates are not substantial compared to the major-group-specific graduation payoffs, consistent with previous finding on limited returns to double or triple majoring (Hemelt, 2010; Rossi and Hersch, 2016).

Yet pre-college interests in major groups seem to play a larger role than learned match quality. The estimates show students' pre-college interests give larger immediate course-taking payoff than match quality beliefs. Although students' pre-college interests have an ambiguous interpretation (they could come family-related influences, or other factors before college that result in overconfidence), they provide potentially strong policy levers to influence major choice.

Students' ACT and SAT Reading and Math Percentiles, and their reported ethnicities and gender affect their course-taking through their match quality beliefs. Estimates of students' characteristics on their prior match quality belief, b_{i1} are shown in Panel D. Reading Percentiles are most influential in the Humanities, and Math Percentiles are most influential in the Natural Sciences.

I find a small gender gap in students' prior beliefs of match qualities. Among Black or Hispanic students, females have slightly lower prior beliefs across all major groups. Among Asian and White students, females have lower prior beliefs in the Social Sciences. However, these gaps are not substantial, with other variables such as Reading Percentiles, Math Percentiles, and high school GPA being more influential.

Panel E in Table 1.4 shows the estimated covariance of major group match qualities. I find positive covariances between all major groups, with Humanities ability being strongly related to the Natural Sciences and Social Science. Consistent with previous findings (Stinebrickner and Stinebrickner, 2014b), the covariance between Natural Sciences and Social Sciences is small. The covariances between Humanities and other major groups are surprisingly large.

[37] Arcidiacono (2004) finds negative payoffs in certain majors based on relative math and verbal abilities, where the comparison case is also dropping out.

[38] Table 1.3 shows that around 57% of all observed choices allocate one or two requirement units in both Humanities and Social Sciences. In addition, Panel E of the model estimates shows the distribution of Humanities and Social Sciences abilities is more narrow, so negative $\kappa_{b,m}$ estimates are identified from smaller variation.

This suggests the Humanities match quality could proxy for a general ability applicable across all major groups. Another reason is the bias that comes from ignoring grades from courses that do not satisfy any major groups' course requirements.

Altogether, the estimates show students with higher Natural Sciences match quality beliefs are more likely to allocate requirement units into the Natural Sciences, and less likely to allocate two requirement units in the Humanities and Social Sciences. This behavior is consistent with students wanting to keep Social Sciences and Humanities as a back-up option in case they receive low grades. Humanities and Social Sciences can be back-up options because while Humanities grades are noisier than Social Sciences grades, the Humanities graduation payoff is more sensitive to GPA.

1.6.2 Evidence of Model Fit

Figure 1.2 shows that the model reasonably fits the external margin of whether students allocate any requirement units in the three major groups, internally transfer, or drop out over time. This fit comes from comparing the actual decisions of a random subset of 16,000 students not used for estimation to the simulated decisions for this random subset across twenty simulations.[39]

The model under-estimates the share of Natural Science graduates, and over-estimates the share of Humanities and Social Science graduates.[40] Figure 1.3 shows that while the estimated model matches when students graduate or otherwise leave CALS, it does not matching the final share of students who graduate in either the Natural Sciences or Social Sciences. Figure 1.4 shows that where around 27% and 47% of students graduate in at least the Natural Sciences and Social Sciences respectively in the data, the model estimates 15% and 53%. The model over-estimates the share of students who double-major in the Humanities and Social Sciences, which drives over estimating the share of students who graduate in at least the Social Sciences.

Figure A.5 shows the model is unable to capture how observed grades are skewed to the left. The estimation procedure tries to fit a normal distribution (from the normality assumption of η_{imt}), and the model over-estimates the share of students who earn the top grade (4.0 out of 4.0). This is most evidence in the Humanities and Social Sciences grades. The skewed pattern is consistent across years.

[39]Figures A.4 and A.5 show the actual decisions – sixteen different requirement unit allocations, internal transfers, and drop out – and grades used for maximum likelihood estimation.

[40]The model does not allow for unobserved heterogeneity through unobserved types. Doing so would increase model fit (Heckman and Singer, 1984; Keane and Wolpin, 1997). I use student characteristics, including pre-college achievement and reported parental educational level, which capture additional variation. Aguirregabira and Mira (2010) provides a comprehensive review. Table A.11 shows that the estimated model reasonably matches the characteristics of students who at least graduate in one of the three major groups, internally transfer, and drop out.

1.6.3 Counterfactual Showing Learning About Match Quality is Important

The first step to decomposing the mechanisms of learning about major match quality and completing majors' course requirements is to test for the presence of learning at the University. Following Stinebrickner and Stinebrickner (2014b), I simulate students' graduating major groups if they did not learn about their major match qualities. In this "no-learning" counterfactual, grades only influence students' graduation payoffs. Results from the "no-learning" counterfactual are consistent with students being initially over confident about their Natural Sciences match quality. I find that this counterfactual increases the share of students who graduate in at least the Natural Sciences by 7 percentage points, from 16% to 23%. Table 1.5 shows that 4 of the 7 percentage point increase comes from the increase in students who only graduate in the Natural Sciences. The remaining 3 percentage points is an increase in students who double-major in the Natural and Social Sciences.

More students will also graduate in other major groups rather than drop out. The drop out share decreases 2 percentage points from 14% to 12%. If students did not learn about their major group match qualities over time, then the share of students who double major in Humanities and Social Sciences would decrease 8 percentage points, from 16% to 8%. The share of students who only graduate in the Social Sciences decreases 4 percentage points, from 24% to 20%.

1.6.4 Counterfactuals Decomposing the Mechanisms of Imposed First-Year Course-Taking

I use two counterfactuals of policy interest to decompose the effects of learned major match quality and completed course requirements on major choice. One way to help students learn about their major group match qualities is to require them to take a diverse set of courses across major groups. To maximize counterfactual information, these courses would be those designed for majors and be required students during their first year. Table A.10 shows 60% of students in the model already choose to allocate one requirement unit in each of the three major groups in their first years.

Table 1.5 shows that when students are required to allocate one requirement unit in each of the three major groups (corresponding to eight to ten courses), the share of students who graduate in at least the Natural Sciences increases 10 percentage points, from 16% to 26%. This increases across all Natural Sciences outcomes: students who only graduate in the Natural Sciences, and graduated in additional major groups. Although the share of students who graduate in the Humanities and Social Sciences decreases 4 percentage points, the overall share of students who graduate in multiple major groups increases.

Two mechanisms drives this result: counterfactual information about major group match quality, and counterfactual completion of major groups' course requirements. Identifying the driving mechanism is crucial, and speaks to how induced course-taking better informs major choice. I adjust whether students learn from their imposed first-year course regime to separate the effects of learning about major match quality from completing course requirements on major choice. This final counterfactual only differs from the penultimate counterfactual in that students believe first year grades are not informative of their major group match qualities.

When students do not learn from their imposed first year of course-taking, I find similar changes in graduation outcomes,[41] suggesting that the driving mechanism is completing course requirements. This is jarring since policy-makers would likely introduce such a policy in order to better inform students' major choices.

Why is the impact on Natural Science graduates so large? The large negative estimate for κ_{2m} in Natural Sciences shows students are essentially unwilling to allocate two requirement units into the Natural Sciences in one year.[42] Therefore, it is difficult for students to "catch up" and graduate in the Natural Sciences by the end of their fourth year if they did not allocate any requirement units into the Natural Sciences in their first year. Therefore the second and third counterfactuals substantially increase the value of continuing to allocate requirement units into the Natural Sciences.

1.7 Discussion and Conclusion

There is public attention on increasing the share of Science, Technology, Engineering, and Math (STEM) majors [43]. Policy-makers or college administrators may try to alter major choice by assigning freshmen students to take a variety of courses, with the intention of better informing their major choices. I show that this more restrictive and information-intensive version of many colleges' distributional course requirements does affect major choice. Combining a dynamic course-taking model with transcript data from a four-year public flagship institution, I find the share of students graduating in the Natural Sciences increases around ten percentage points, with the Humanities or Social Sciences decreasing around two percentage points.

Where previous literature has confounded the effects of learning about major match quality and completing majors' course requirements on major choice, this paper finds the effect of completing majors' course requirements looms larger. Although students have additional information about their major group match qualities, they had to complete course requirements to get it. Completing these course requirements decreases the effort needed to graduate and

[41]Table A.12 suggests that the counterfactual students in these graduating outcomes are not substantially different. It shows that there are no substantial differences in the characteristics of students between these two counterfactuals.

[42]Table 1.3 shows that in 1% of all observed course-taking, students complete 50% of the course requirements for the Natural Sciences in one year.

[43]former-President Barack Obama, https://obamawhitehouse.archives.gov/blog/2016/02/11/stem-all; Homeland Security Department, Improving and Expanding Training Opportunities for F-1 Nonimmigrant Students With STEM Degrees and Cap-Gap Relief for All Eligible F-1 Students; ComputerWeekly.com, http://www.computerweekly.com/news/4500254034/IT-industry-calls-for-government-to-fill-Stem-skills-gap

drives students course-taking trajectories.

Though the policy is intended to inform students about their major match quality, it instead sets students on a path towards graduating in different majors. Students are responding to being closer to finishing the course requirements in the Natural Sciences, rather than having more information about their major match quality in the Natural Sciences.

Key to understanding the effects of induced course-taking is how students go from initial interests and beliefs about major match quality to their graduating majors. Not only do students want to learn about their major match qualities, but also need to consider whether to respond to learned match qualities. This paper provides a theoretical framework for how students may ignore learned major match quality because they are not willing to take the courses needed to graduate in that major. This paper starts to close the literature's gap around the effort needed to graduate in different majors.

Table 1.1: Aggregated Major Groups

Percentage of Graduates Shown in Parenthesis for Sample of Students

Major Groups in the College of Arts and Liberal Sciences (CALS):

- Natural Sciences
  - Biology (18.15%)
  - Physics (7.18%)
  - Earth Sciences (2.02%)
  - Mathematics or Statistics (1.71%)
  - Computer Science (1.41%)
- Humanities
  - Foreign Languages (1.29%)
  - Area Studies (3.12%)
  - History and Art Hist. (1.58%)
  - Philosophy (1.11%)
  - English and Creative Writing and Literature (1.25%)
  - Communications (1.49%)
- Social Sciences
  - Economics (6.48%)
  - Psychology (7.32%)
  - Sociology (2.62%)
  - Political Science (2.77%)
  - Anthropology (0.59%)

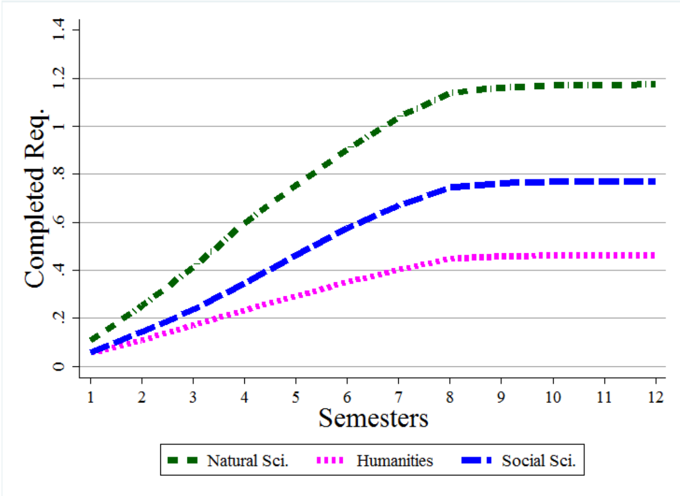
Internal Transfer Options to Colleges at the University:

- Business
  - Business (8.79%)
- Other
  - Engineering (2.23%)
  - Medical-Related (2.36%)
  - Public Policy (1.00%)
  - Art-Related (1.98%)
  - Education (1.75%)

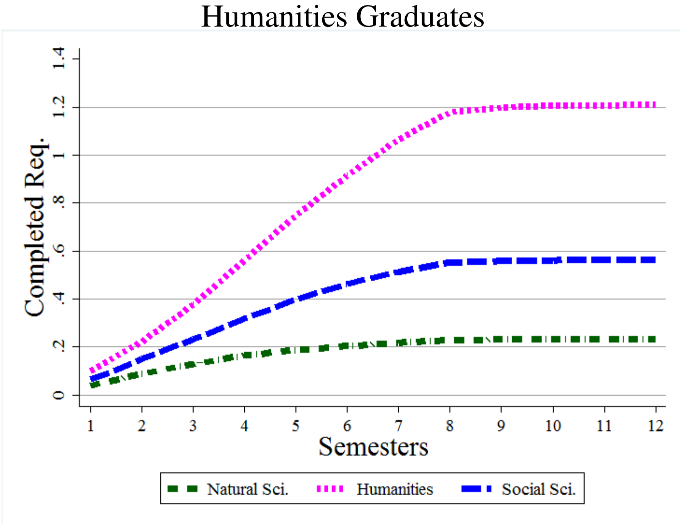
NOTES – The table does not show extensive lists. Instead it shows the most common individual majors within each major group.

Figure 1.1: How Different Graduates Fulfill All Major Groups' Course Requirements Over Time

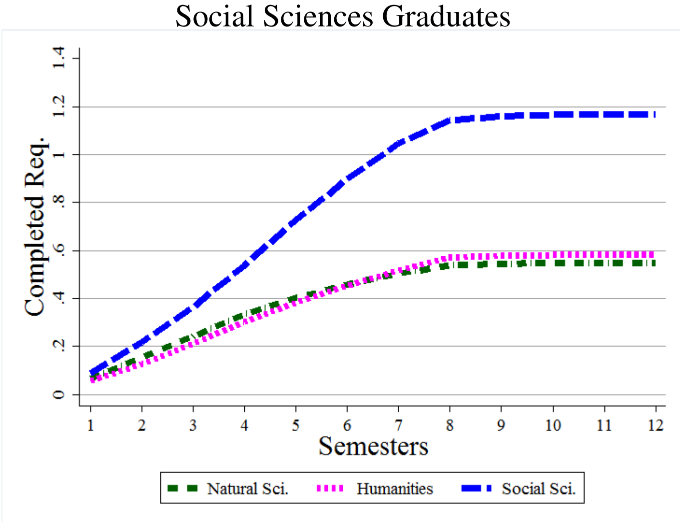
Natural Sciences Graduates



Humanities Graduates



Social Sciences Graduates



NOTES – Cumulative progress is calculated using the proportion of each major groups' course requirement that are satisfied at the end of each semester. Progress within a major group comes from taking the maximum progress of individual majors within a major group.

Table 1.2: Sample Selection Table: Empirical and Model-Driven Concerns

	Start	Dropped Due to... Empirical Concerns	Model Concerns	Final
Time-Invariant Characteristics:				
Female	0.563	0.482	0.576	0.574
Black	0.056	0.010	0.097	0.062
Asian	0.138	0.235	0.107	0.126
Hisp	0.052	0.025	0.066	0.055
Reading Perc.	83.621	90.018	81.279	82.811
Math Perc.	86.222	92.776	81.618	85.447
AP Credits	7.237	21.439	5.009	5.362
Total Transfer Credits	2.946	6.644	2.328	2.458
International (non-US Citizen)	0.056	0.063	0.057	0.055
Self-Reported Annual Household Income				
$<$ 100K	0.213	0.148	0.270	0.221
$\geq$ 100K	0.401	0.421	0.320	0.400
Missing	0.386	0.432	0.410	0.379
Self-Reported Maximum Parental Education				
HS Graduate	0.079	0.056	0.106	0.082
College	0.284	0.205	0.273	0.295
Advanced Degree	0.451	0.527	0.396	0.442
Missing	0.186	0.213	0.225	0.181
Pre-College Interest				
Natural Sciences	0.435	0.528	0.399	0.423
Humanities	0.359	0.373	0.434	0.356
Social Sciences	0.357	0.389	0.297	0.354
Other College	0.367	0.358	0.370	0.368
Business College	0.271	0.282	0.175	0.272
Outcomes – Graduation, Internal Transfer, and Dropout:				
Internal Transfer – Other College	0.099	0.130	0.068	0.096
Internal Transfer – Business College	0.084	0.167	0.000	0.075
Drop Out	0.101	0.088	0.300	0.098
Graduation from CALS	0.716	0.615	0.632	0.731
N	41331	4851	691	35789

NOTES – Shown Reading and Math percentiles are students' ACT and SAT last reported scores represented as percentiles. If a student has both ACT and SAT scores, then the percentiles are averaged. Self-reported household income and maximum parental education comes from the Common Application. Advanced degrees include Masters, Doctorates, and other advanced professional degrees. Students' citizenship is collected at application time. Transfer credits are all credits from other institutions that are earned before students start any coursework at the University. Table A.9 breaks down the Empirical and Model Concerns in detail.

Table 1.3: Distribution of Requirement Unit Allocations over Years, Final Sample

Requirement Unit Allocation	Years: Number of Students 1	2	3	4	Yearly Shares 1	2	3	4
0 0 0	390	325	453	743	0.011	0.010	0.016	0.030
0 0 1	1292	1180	1695	1977	0.036	0.037	0.060	0.079
0 0 2	23	89	168	115	0.001	0.003	0.006	0.005
0 1 0	1888	2236	3099	3944	0.052	0.071	0.109	0.157
0 1 1	5285	7183	8666	7093	0.147	0.227	0.305	0.282
0 1 2	99	268	183	145	0.003	0.008	0.006	0.006
0 2 0	76	128	227	217	0.002	0.004	0.008	0.009
0 2 1	355	239	196	179	0.010	0.008	0.007	0.007
1 0 0	1473	842	507	566	0.041	0.027	0.018	0.023
1 0 1	4767	3723	3635	2584	0.132	0.117	0.128	0.103
1 0 2	22	139	156	124	0.001	0.004	0.005	0.005
1 1 0	4410	3241	2295	1968	0.122	0.102	0.081	0.078
1 1 1	15619	11895	7013	5258	0.434	0.375	0.247	0.209
1 2 0	267	109	77	83	0.007	0.003	0.003	0.003
2 0 0	7	8	13	32	0.000	0.000	0.000	0.001
2 0 1	22	41	16	30	0.001	0.001	0.001	0.001
2 1 0	23	40	20	81	0.001	0.001	0.001	0.003

NOTES –

Starting from the left, requirement units are allocated in the Natural Sciences, Humanities, and Social Sciences.

Table 1.4: Estimates from the Dynamic Course-Taking Model

Panel A: **Immediate Payoffs from Course-Taking, found in (1.3) and (1.4)**

$$u_{imt} = \kappa_{1m}\mathbb{1}\{\mathbb{c}_{imt}=1\} + \kappa_{2m}\mathbb{1}\{\mathbb{c}_{imt}=2\} + \kappa_{\text{adv},m}\,\text{freqadv}(\mathbb{c}_{imt}) + \mathbb{c}_{imt}\Big(\kappa_{b,m}b_{imt} + \kappa_{z,m}Z_i\Big)$$

$$u(\mathbb{c}) = \pi_1\mathbb{1}\{\tilde{\mathbb{c}_{it}}=1\} + \pi_2\mathbb{1}\{\tilde{\mathbb{c}_{it}}>1\} + \pi_{var}\,\text{variance}(\mathbb{c}) + \epsilon_{i\mathbb{c}t}$$

where $\tilde{\mathbb{c}_{it}} = \sum_{m'=1}^{M}\mathbb{1}\{e_{im't}>0\}$, and $\epsilon_{i\mathbb{c}t} \sim EV(0,\tau_t)$

	Natural Sciences	Humanities	Social Sciences
κ_{1m}	–1.112∗∗∗	1.686∗∗∗	1.561∗∗∗
	[0.185]	[0.194]	[0.222]
κ_{2m}	–6.221∗∗∗	0.52	–0.162
	[0.355]	[0.335]	[0.399]
$\kappa_{\text{adv},m}$	0.519∗∗∗	0.617∗∗∗	1.132∗∗∗
	[0.121]	[0.137]	[0.152]
$\kappa_{b,m}$	0.166∗∗∗	–0.345∗∗∗	–0.308∗∗∗
	[0.031]	[0.032]	[0.042]
$\kappa_{\text{Pre-College Interest},m}$	0.713∗∗∗	0.251∗∗∗	0.177∗∗∗
	[0.032]	[0.028]	[0.032]
$\kappa_{\text{Non-US Citizen},m}$	0.367∗∗∗	–0.458∗∗∗	0.048
	[0.097]	[0.095]	[0.098]
$\kappa_{z,m}$ on *Reported Annual Household Income:*			
$\kappa_{z,m} \geq$ \$100K	–0.035	–0.064	0.007
	[0.053]	[0.050]	[0.055]
$\kappa_{z,m}$ is missing	–0.045	–0.002	–0.064
	[0.059]	[0.057]	[0.062]
$\kappa_{z,m}$ on *Maximum Parental Education:*			
$\kappa_{z,m}$ College	0.041	–0.308∗∗∗	–0.175∗∗
	[0.077]	[0.071]	[0.085]
$\kappa_{z,m}$ Advanced Degree	0.038	–0.189∗∗∗	–0.158∗
	[0.077]	[0.070]	[0.083]
$\kappa_{z,m}$ Missing	0.012	–0.106	–0.286∗∗∗
	[0.094]	[0.089]	[0.099]

π_1	–4.705∗∗∗	τ_2	1.59∗∗∗
	[0.308]		[0.05]
π_2	–4.261∗∗∗	τ_3	2.00∗∗∗
	[0.429]		[0.08]
π_{variance}	–2.696∗∗∗	τ_4	2.59∗∗∗
	[0.661]		[0.12]
		$\varphi_{\mathbb{L}}$	0.75∗∗∗
			[0.12]

Table 1.4 Continued, Labor Market and Internal Transfer Payoffs

Panel B: Labor Market Payoffs, found in (1.6) and (1.7)

$$W_{imt} = \rho_{m,t=3} + \rho_{m,t=4} + \rho_{g,m}\overline{g_{imt}} + \rho_{1m}\,\text{2006to2007Cohort}_i + \rho_{2m}\,\text{2008to2011Cohort}_i$$

$$\widetilde{W}_{i\ \text{NatSci \& Human}\ t} = \tfrac{1}{2}W_{i\ \text{NatSci}\ t} + \tfrac{1}{2}W_{i\ \text{Human}\ t} + \rho_{\text{NatSci \& Human}}$$
$$\widetilde{W}_{i\ \text{NatSci \& SocSci}\ t} = \tfrac{1}{2}W_{i\ \text{NatSci}\ t} + \tfrac{1}{2}W_{i\ \text{SocSci}\ t} + \rho_{\text{NatSci \& SocSci}}$$
$$\widetilde{W}_{i\ \text{Human \& SocSci}\ t} = \tfrac{1}{2}W_{i\ \text{Human}\ t} + \tfrac{1}{2}W_{i\ \text{SocSci}\ t} + \rho_{\text{Human \& SocSci}}$$
$$\widetilde{W}_{i\ \text{NatSci \& Human \& SocSci}\ t} = \tfrac{1}{3}W_{i\ \text{NatSci}\ t} + \tfrac{1}{3}W_{i\ \text{Human}\ t} + \tfrac{1}{3}W_{i\ \text{SocSci}\ t} + \rho_{\text{NatSci \& Human \& SocSci}}$$

	Natural Sciences	Humanities	Social Sciences	Multiple Major Groups	
ρ_{1m}	–0.498	–0.283	–0.491*	$\rho_{\text{NatSci \& Human}}$	0.000
	[0.372]	[0.297]	[0.283]		[0.002]
ρ_{2m}	–1.039***	–0.838***	–0.404*	$\rho_{\text{NatSci \& SocSci}}$	0.024***
	[0.331]	[0.255]	[0.242]		[0.001]
$\rho_{m,t=3}$	4.563***	3.643***	5.939***	$\rho_{\text{Human \& SocSci}}$	0.000
	[0.866]	[0.782]	[0.705]		[0.002]
$\rho_{m,t=4}$	4.363***	3.083***	4.077***	$\rho_{\text{NatSci \& Human \& SocSci}}$	0.002*
	[0.771]	[0.760]	[0.684]		[0.001]
$\rho_{g,m}$	1.521***	1.541***	0.911***		
	[0.237]	[0.221]	[0.192]		

Panel C: Business and Other Internal Transfer Payoffs, found in (1.5)

$$u^{Bus.} = \kappa^{\text{Bus.}}_{t=2} + \kappa^{\text{Bus.}}_{t=3} + \kappa^{\text{Bus.}}_{t=4} + \kappa^{\text{Bus.}}_1\overline{g_{i,\ \text{NatSci},\ t}} + \kappa^{\text{Bus.}}_2\overline{g_{i,\ \text{Human},\ t}} + \kappa^{\text{Bus.}}_3\overline{g_{i,\ \text{SocSci},\ t}} + \kappa^{\text{Bus.}}_{\text{int}}\text{Pre-College Interest in Bus.}_i + \epsilon_{i\mathfrak{c}t}$$
$$u^{Oth.} = \kappa^{\text{Oth.}}_{t=2} + \kappa^{\text{Oth.}}_{t=3} + \kappa^{\text{Oth.}}_{t=4} + \kappa^{\text{Oth.}}_1\overline{g_{i,Nat.Sci.,t}} + \kappa^{\text{Oth.}}_2\overline{g_{i,\ \text{Human},\ t}} + \kappa^{\text{Oth.}}_3\overline{g_{i,\ \text{SocSci},\ t}} + \kappa^{\text{Oth.}}_{\text{int}}\text{Pre-College Interest in Oth.}_i + \epsilon_{i\mathfrak{c}t}$$

Business	
$\kappa^{\text{Bus.}}_{t=2}$	–1.067**
	[0.522]
$\kappa^{\text{Bus.}}_{t=3}$	–2.689***
	[0.866]
$\kappa^{\text{Bus.}}_{t=4}$	–8.525***
	[1.638]
$\kappa^{\text{Bus.}}_1$	1.347***
	[0.332]
$\kappa^{\text{Bus.}}_2$	–0.493**
	[0.223]
$\kappa^{\text{Bus.}}_3$	0.274
	[0.401]
$\kappa^{\text{Bus.}}_{\text{int}}$	3.308***
	[0.280]

Oth.	
$\kappa^{\text{Oth.}}_{t=2}$	–0.626*
	[0.320]
$\kappa^{\text{Oth.}}_{t=3}$	0.226
	[0.515]
$\kappa^{\text{Oth.}}_{t=4}$	–3.950***
	[0.838]
$\kappa^{\text{Oth.}}_1$	1.734***
	[0.179]
$\kappa^{\text{Oth.}}_2$	–0.490***
	[0.186]
$\kappa^{\text{Oth.}}_3$	–0.508***
	[0.179]
$\kappa^{\text{Oth.}}_{\text{int}}$	1.330***
	[0.214]

Table 1.4 Continued

Panel D: **Prior Beliefs, found in (1.9)**

$$b_{im1} = \phi_m^o + \phi_m X_i + \epsilon_{im}^b$$
$$g_{imt} = \mu_{im} + \eta_{imt}$$
$$\eta_{imt} \sim \mathrm{N}(0, \sigma_m^2)$$

	Natural Sciences	Humanities	Social Sciences
ϕ_m Reading Percentile	0.000	0.003∗∗∗	0.002∗∗∗
	[0.001]	[0.001]	[0.001]
ϕ_m Reading Percentile 80-90	0.066∗∗	0.110∗∗∗	0.067∗∗
	[0.031]	[0.032]	[0.030]
ϕ_m Reading Percentile $\geq$ 90	0.116∗∗∗	0.176∗∗∗	0.132∗∗∗
	[0.039]	[0.040]	[0.036]
ϕ_m Math Percentile	0.018∗∗∗	0.001	0.007∗∗∗
	[0.001]	[0.001]	[0.001]
ϕ_m Math Percentile 80-90	–0.039	0.060∗	–0.004
	[0.034]	[0.034]	[0.031]
ϕ_m Math Percentile $\geq$ 90	0.109∗∗∗	0.083∗∗	0.046
	[0.041]	[0.038]	[0.035]
ϕ_m High School GPA	1.062∗∗∗	0.592∗∗∗	0.655∗∗∗
	[0.030]	[0.025]	[0.027]
ϕ_m High School GPA $\geq$ 3.5	–0.073∗∗∗	0.152∗∗∗	–0.01
	[0.031]	[0.029]	[0.028]
ϕ_m Black_Hisp	–0.043	–0.182∗∗∗	–0.093∗∗
	[0.047]	[0.036]	[0.041]
ϕ_m Black_Hisp × F	0.015	0.031	–0.069
	[0.061]	[0.047]	[0.051]
ϕ_m Asian	–0.005	–0.188∗∗∗	–0.139∗∗∗
	[0.054]	[0.055]	[0.050]
ϕ_m Asian × F	–0.076	0.145∗	0.002
	[0.071]	[0.076]	[0.067]
ϕ_m Female	–0.060∗∗∗	–0.058∗∗∗	0.072∗∗∗
	[0.023]	[0.024]	[0.022]
ϕ_m^o	–2.982∗∗∗	0.358∗∗∗	–0.428∗∗∗
	[0.055]	[0.052]	[0.057]
σ_m^2	0.843∗∗∗	0.875∗∗∗	0.733∗∗∗
	[0.009]	[0.006]	[0.006]

Table 1.4 Continued

Panel E: **Covariance of Match Qualities, found in (1.10)**

$$\mathbb{E}_t[\mu_i] = b_{it} = \mathbb{V}_t[\mu_{it}] \cdot \Big((\Delta)^{-1} b_{i1} + (\Sigma)^{-1} \sum_{t'=1}^{t} g_{it'}\Big)$$

$$\mathbb{V}_t[\mu_i] = \Big((\Delta)^{-1} + (\Sigma)^{-1} (\sum_{t'=1}^{t} \mathbb{c}_{it})\Big)^{-1}$$

where the matrix Σ only has diagonal entries σ_m^2

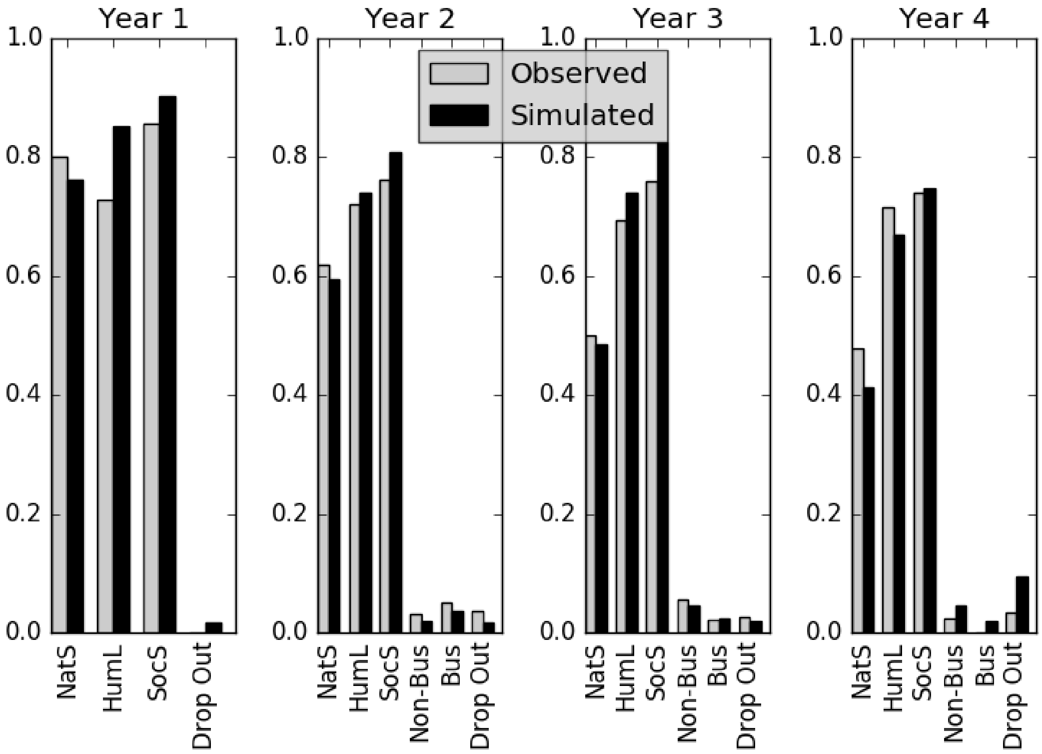
Δ	Natural Sciences	Humanities	Social Sciences
Natural Sciences	4.114∗∗∗	—	—
	[0.135]	—	—
Humanities	6.118∗∗∗	1.563∗∗∗	—
	[0.104]	[0.052]	—
Social Sciences	3.137∗∗∗	4.634∗∗∗	2.473∗∗∗
	[0.091]	[0.84]	[0.083]

Standard errors in brackets. $^{***} - p < 0.01,^{***} - p < 0.05,^{***} - p < 0.10$

NOTES – Parameter estimates on Reported Household Income in **Panel A** are relative to students who report annual household income less than $100,000. Parameters on Maximum Parental Education are relative to students who report a high school degree or less. A maximum parental education of college includes attending "some college," associates, and bachelors degree. "Some college" and associates degrees have very small shares. Advanced Degrees includes masters, doctorates, medical, and legal degrees.

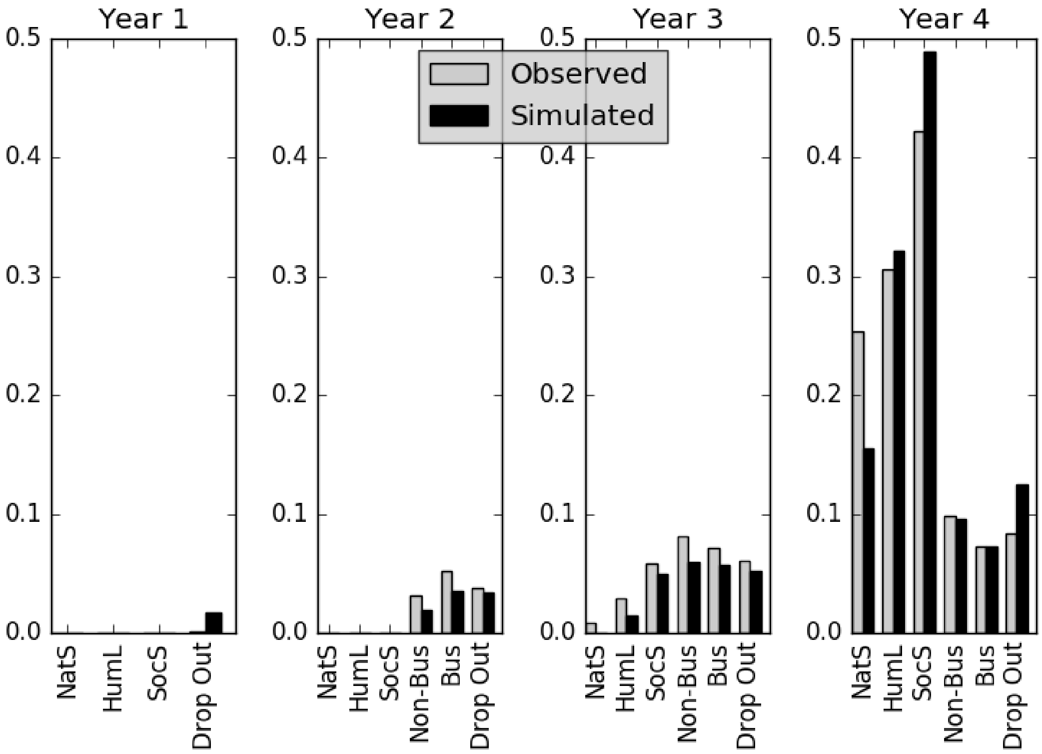
Parameters are estimated with maximum likelihood estimation, where convergence comes from the Limited Memory Broyden–Fletcher–Goldfarb–Shanno Algorithm. I restrict σ_m^2, τ_t, and diagonal elements of Δ to be strictly positive, $\varphi_{\mathbb{L}}$ to be within 0 and 1, and Δ to be symmetric. Corresponding standard errors are calculated with the delta method. I estimate using a random subset of 4,000 students from the final estimation sample.

Figure 1.2: Model Fit: Share of Allocating Any Requirement Units Across Major Groups, Internal Transfers, and Drop Out



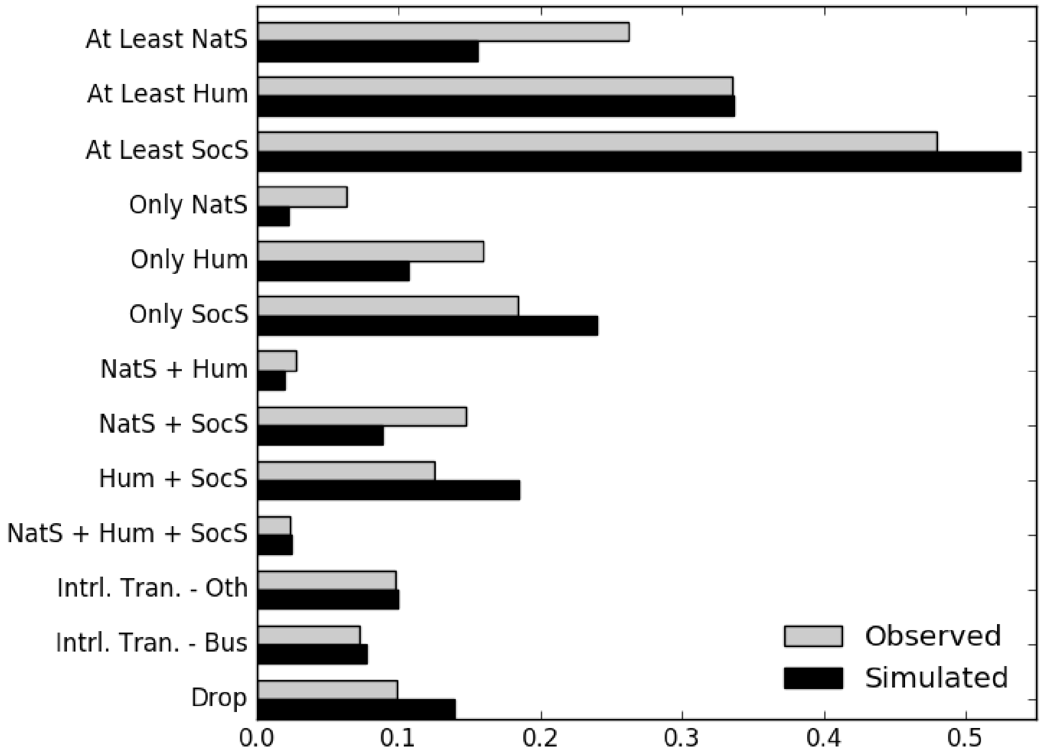
NOTES – This compares the share of students who make any progress in different major groups, internally transfer out of CALS, and drop out. Simulated shares come from twenty simulations, with draws of ϵ_{ict} and η_{imt}. I run 20 simulations on a random 16,000 students in the non-estimation sample.

Figure 1.3: Model Fit: Share of Graduation Outcomes Over Time



NOTES – This compares the share of students who graduate in any of the major groups over time, internally transfer out of CALS, and drop out. It is possible to graduate in multiple majors, if the student finishes all requirements in the same year. Since students graduate in multiple majors, the shares do not add up to one. Simulated shares come from twenty simulations, with draws of $\epsilon_{i\overline{c}t}$ and η_{imt}. I run 20 simulations on a random 16,000 students in the non-estimation sample.

Figure 1.4: Model Fit: Share of Graduation Outcomes



NOTES – This shows the cumulative share of students who graduate in mutually exclusive combinations of major groups, internally transfer, or drop out.

Table 1.5: Counterfactual Results: Share of Graduates and Other Outcomes

	Baseline Simulation	Shares from Counterfactual Simulations: (1) "No-Learning"	(2) First Year Course Regime that is Informative of Match Quality: Yes	(3) First Year Course Regime that is Informative of Match Quality: No
Share of Students Over Detailed CALS Graduation Outcomes				
Only Natural Sciences	0.023	0.063	0.054	0.057
Only Humanities	0.107	0.152	0.143	0.142
Only Social Sciences	0.240	0.200	0.190	0.190
Natural Sciences & Humanities	0.020	0.029	0.037	0.037
Natural Sciences & Social Sciences	0.089	0.119	0.129	0.140
Humanities & Social Sciences	0.185	0.104	0.149	0.140
All Major Groups	0.024	0.021	0.038	0.037
Share of Students in Overall Graduation Outcomes				
Natural Sciences	0.156	0.233	0.258	0.271
Humanities	0.336	0.307	0.367	0.357
Social Sciences	0.539	0.446	0.506	0.506
Other College	0.100	0.104	0.083	0.088
Business College	0.077	0.086	0.069	0.076
Drop Out	0.139	0.120	0.107	0.092

NOTES – In (1), students believe grades are not informative of their major group match qualities and do not update their beliefs over time. In (2) and (3), students are required to satisfy 25% of the course requirements in each major group in their first year. First year grades inform students about their match quality beliefs in (2), but not (3). Results are from twenty simulations on a random subset of 16,000 students from the final sample not used for estimation.

Chapter 2

Do Grades Matter? Evidence from College Transcripts

2.1 Introduction

Despite strong evidence of large labor market returns to college major (Altonji et al., 2015; Hastings et al., 2013; Kirkebøen et al., 2016; Varnevale et al., 2013), there is evidence students' major choices are not fully informed, and would benefit from learning about their capacity to perform in academic and non-academic environments related to majors. While this does not seem surprising considering the limited exposure high school students have to different majors (Gottfried and Bozick, 2016), it presents an abundance of room for policy to inform major choice.

The previous literature has examined how students learn about their fit, or ability, in a major from the one grade they earn after declaring in it. A vital and arguably understudied component of major choice is that majors' course requirements separate major choice into a sequence of choices. Courses, not major declarations, are the information vector to learn about major ability. Then it is vital to consider how students are receiving grades from many majors at a time. This paper examines the extent to which course grades influence students' decisions to complete majors' course requirements.

This paper makes three contributions to the literature. First, it is among the first to use college transcript data to understand how students take the courses needed to graduate in different majors. Rather than major declarations, these courses measure the necessary steps to graduating in a major and track how students interact with multiple majors at a time. Using administrative transcript data from a public four-year institution, I identify the courses needed to graduate in different majors. This approach cannot be done with nationally representative datasets such as the National Longitudinal Survey of Youths or National Education Longitudinal Study because they do not have sufficient sample size to make comparisons between majors within an institution.

Second, where previous works use major declarations to proxy for students' information sets and intentions to graduate in a major (Arcidiacono, 2004), this paper finds diverse course-taking across majors. Students take courses across majors before declaring a major, suggesting their major choices are informed. After declaring a major, students continue to take courses needed to graduate in the Social Sciences, Humanities, and Psychology majors. Finding diverse course-taking persists after declaration argues major declarations are not sufficient to measure how students engage different majors.

The third contribution is a single-agent dynamic course-taking model that features how students take relevant courses to learn about their major abilities and complete majors' course requirements. The model conceptually defines ability as students' performance-related match quality. Focusing on course-taking rather than major declarations, the model provides a framework for understanding how students may diversify their course-taking relevant to majors in case they learn they have low major ability, or focus their course-taking in a major if they are confident in their ability and want to hasten graduation.

The course-taking model's comparative statics provide a framework to understand how student course-taking respond to earned grades. Students who receive higher grades are more likely to continue taking relevant courses in a major. Course-taking history within and across majors are also substantial, suggesting students persist in a major independent of ability beliefs. This is likely due to a desire to graduate quickly, and is a substantial hurdle for successfully influencing student major choice during college.

The second section discusses how the paper fits into the current college major and course choice literature. The third section discusses the administrative transcript data and presents descriptive trends in course-taking. The fifth and sixth sections describes the estimation strategy and results. The sixth section discusses and concludes with how transcript data can be used in future work.

2.2 Literature Review

This paper is among the first to look at college transcript data to understand college major choice and graduation outcomes. Focusing on college majors, it is tied to the college major choice literature. Although a branch of the college major choice literature finds that labor market returns[1] play a significant role (Beffy et al., 2012; Long et al., 2014; Montmarquette et al., 2002), a separate branch has looked at the role of non-pecuniary factors. Students often do not know their own abilities, and the literature has found students changing majors in response to low grades (Arcidiacono et al., 2012a; Stinebrickner and Stinebrickner, 2014a,b). These abilities affect earnings and payoffs during college. Payoffs during college come from the

[1]Students often list labor market returns as an important factor for their college major choice (Malgwi et al., 2005). This is particularly interesting in light of evidence that students are often surprised to learn the earnings of majors they have not declared in (Arcidiacono et al., 2012b).

effort exerted, grades earned, and other psychic values of being in different majors.

Previous approaches assume that students only receive grades in their declared major, and are ambiguous on how all students transition from being "undeclared" to being in different majors. This paper uses transcript data to study how students are receiving grades across multiple majors at a time. Looking at courses not only captures how students engage in different majors, but also measures the effort needed to finish.[2] Bettinger (2010) also uses transcript data from the University System of Ohio, although it does not use college grades.

Ost (2010) uses a regression discontinuity design to find insubstantial impacts of letter grade assignments in an introductory microeconomics course on majoring in Economics and continuing to take courses in Economics, at a selective research university. Owen (2008) uses the same design and finds positive impacts on females in introductory microeconomics courses, in a different selective research university and liberal arts college. While these works use regression discontinuity to study these effects in isolation from other course-taking, this paper embraces the multi-faceted way students are receiving multiple grades.

To fully engage the transcript data, it is vital to track how students progress through majors' course requirements. This relies on identifying the courses needed for individual majors, identifying the most relevant courses for each major. To my knowledge, the college course literature has not exploited institutional details to examine course-taking in different majors over time.

2.3 Student Transcript Data and Major Progress

I use administrative student transcript data from a large public four-year institution, anonymized as the Public Higher Education Consortium (PHEC). This transcript data records each course students take at PHEC, as well as each course credit transferred in from outside institutions, and Advanced Placement and International Baccalaureate exams.

The largest school at PHEC is the "College of Arts and Sciences" (CAS), with over 60% of total PHEC enrollment. The data universe is all students who enroll at PHEC through CAS from 2002 to 2012. While students are not required to start in CAS, CAS offers the most diverse set of majors, and students interested in majors in business, architecture, public policy, and other medical related fields must transfer out of CAS. This internal transferring process has a separate application progress, and offers different tuition rates. Most of these other schools at PHEC require students to take specific courses across PHEC to have any consideration. I observe student course-taking after they internally transfer out of CAS.

Other administrative data includes reported sex, ethnicity, and zip code at the time of admission. PHEC collects students' SAT and ACT scores. I use percentile scores provided by

[2]Hendricks and Leukina (2015) uses college transcript data to measure college completion, without focusing on majors. This is largely because it uses data across multiple institutions with different course requirements and focuses on the margin of whether students complete any graduation requirements instead of completing different majors.

the College Board and ACT to normalize these scores. I focus on students' Reading and Math percentiles. Admission records also indicate students' interests in different careers and majors prior to enrolling at PHEC. I also observe when students officially declare and file to graduate in different majors. Students complete majors in this data, not based on their coursework, but when they successfully file for graduation. Students in certain residential programs at PHEC do not declare majors.

To identify course-taking relevant to completing different majors, I code each major's course requirements to track how students explore different majors.[3] Designed for majors, these courses are likely the most informative of students' major match quality. With more than 50 majors at CAS and overlapping course requirements,[4] I aggregate majors into seven majors, shown in Table 2.1. There are two significant deviations from previous works. First, Economics and Psychology are their own majors because they are the two largest majors at CAS. Second, commonly used Business, Engineering, and Education majors (Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2014b) are not present because they are not offered in CAS.

For each of these seven aggregated majors, I measure the amount of progress students have made to completing the course requirements for each major. I use the maximum amount of progress of individual majors within the aggregated majors and within a semester. While this masks the progress students make within a major, it is the relevant statistic[5] for measuring progress towards completing any individual major within that aggregated major group.

One potential issue with interpreting course-taking as interest in majors are general distributional course requirements. At PHEC, students cannot use the same class to satisfy distributional and major requirements, and there are many courses outside general distributional course requirements satisfy major requirements. This is a potential issue because I am using whether students take courses required for majors as an outcome of their grades. Students taking courses required for majors to instead satisfy distributional course requirements attenuates the estimated relationship between grades and course-taking to zero.

To understand the upper bound of measurement error from inferring major interest from course-taking, Figure 2.1 shows that around 70% of all individual first and second semester courses taken satisfy general distributional requirements and at least one major's course re-

[3] If the student earns less than "C–" in the course, it does not satisfy that major's course requirement. I code students' initial attempt to complete each course requirements, and less than 4% of all observed courses have a grade less than "C–." Course withdrawals are only observed if the student withdrew from the course at least three weeks after the semester starts.

[4] The multitude of majors also affects interpretation. Changing from Chemistry to Biology courses is very different from changing from Biochemistry to History major. The latter is of more interest in the model, where the former choice can be interpreted as a change of focus within a larger group.

[5] Another possible measure is to use the progress from the "relevant major." Say that a student has already made 10% progress in Biology and 20% in Chemistry within the Life & Earth Sciences major. If the student makes 10% progress in Biology and 5% progress in Chemistry in the next semester, then I would record the 5% as the "relevant major" it follows the major the student is closest to finishing. However, this under-measures the student's effort in the Life & Earth Sciences major.

quirements. Soon after their second semesters, students quickly concentrate on taking courses that only satisfy majors' course requirements. Starting from students' third semesters, the proportion of courses that only satisfy at least one major's course requirements sharply increases to 60%, and stays at 80% afterwards.

2.3.1 How Different Graduates Complete Majors' Course Requirements

Although all CAS students face the same set of general distributional requirements, there is variation in how they take courses to progress through different majors ' course requirements. From these trends of cumulative progress, I infer how students explore other majors. Since course-taking is not limited by major declarations at PHEC, I interpret differences in course-taking between graduates of different majors as exploring. The variation suggests students explore other majors to varying degrees.

I show students' cumulative progress in satisfying all majors ' course requirements in Figure 2.2, for Economics and Psychology graduates. On average, Economics graduates make 30% to 40% progress across all majors (except Arts) by their fourth semester. Psychology graduates' progress in Psychology has already outstripped their progress in other major groups by their fourth semesters. Compared to Psychology graduates, Economics graduates make progress in more majors – consistent with additional exploring. An alternative explanation is Economics graduates are worried they will meet the majors' minimum grade requirements. However, only 5% of all assigned course grades are below the minimum grade requirement (C-).

By their fourth semesters, Psychology graduates have less diverse course-taking histories than Economics graduates. They have completed around 30% of the course requirements in Life & Earth Sciences, Humanities, and Social Sciences, 20% in Math & Physical Sciences, and 10% in Economics. Compared to Economics graduates, Psychology graduates are not as exposed to other majors.

2.3.2 Course-Major Correspondence

After showing diverse course-taking across majors, how could major declarations be under-measuring how students take courses and participate in other majors? In Table 2.2, I compare students' progress across majors with their declarations.

Students' progress across majors groups the semester before they declare a major suggest how informed their major declarations are. Economics students have made around 25% progress across all majors (excluding Arts). Math & Physical Sciences and Life & Earth Sciences students have made the most amount of progress in their respective majors the semester before they declare, and little progress in other majors. Math & Physical Sciences and Life & Earth Sciences students make less progress in the Humanities than all other students. Humanities and Social Science students behave similarly, with concentrated progress.

Although this is not definite evidence of how informed students' major choices are, students' concentrated progress can make it more difficult to switch majors. Table 2.2 shows students declare their majors after their fourth semester. Humanities and Social Sciences students, by the time they have declared, will on average find it more difficult to switch to Math & Physical Sciences or Life & Earth Sciences majors simply because they have not made as much progress in them.

If policy-makers are interested in influencing major choice, then prior to major declaration is likely a more effective time to do so. Table 2.2 shows that four semesters after declaring a major, students have specialized their course-taking into their declared majors. However, there are still patterns consistent with continued exploration. Students who did not declare in the Humanities, Social Sciences, and Psychology make 15% to 25% progress in them. Potential reasons for continued exploration is students want to learn about their abilities in these majors, desire to graduate in these majors, completing minors in these majors, and the course-taking experience itself.

2.3.3 Evidence of Ability Sorting

If students are learning about their abilities across majors, then there will be variation over the grades they earn. Figure 2.3 suggests students make progress in the Math & Physical Sciences and Humanities majors in response to their first semester GPAs in these majors.

Figure 2.3 shows that students in higher quartiles of first semester Math & Physical Sciences GPA make more progress in this majors than students in lower quartiles. This is consistent with students learning their ability in Math is higher and believing the future payoff of taking Math courses is directly related to their Math & Physical Sciences ability. Figure 2.3 shows the same pattern for the Humanities. The Math & Physical Sciences trend has more variance, implying students are more sensitive to Math & Physical Sciences than Humanities grades.

To help explain why students are taking courses across different majors over time, I develop a dynamic course-taking model where courses provide students with information about their major-specific abilities, and also complete the course requirements needed to graduate in those majors.

2.4 A Model of College Course-Taking

Where the previous literature has used models of major declaration (Arcidiacono, 2004), I develop a dynamic course-taking model to understand how students take courses to learn about their major abilities and complete majors' course requirements. The model can be used to understand the diverse course-taking patterns before and after major declarations in Tables 2.2. Focusing on course-taking provides a new way to interpret course-taking as a continuous measure of the student's revealed preference to graduate in that major and learn about her major abilities. The model provides a framework for how earned grades influence students' course choices, and why students persist in a major because of previous course-taking.

The model resembles Altonji (1993)'s model of how students work towards graduating in different majors. The student takes a bundle of courses across M majors to maximize her future discounted sum of college flow payoffs and graduation payoffs. Courses can only satisfy one major's course requirements, and give the student an immediate flow payoff during college. When the student completes a major's course requirements, she graduates in that major and receives a major-specific graduation payoff. The student may drop out of college at any point. Dropping out and graduating are terminal states of the model.

Tracking progress to completing majors' course requirements makes graduation time endogenous, where in previous models students spent an exogenous amount of time in college (Arcidiacono, 2004; Kinsler and Pavan, 2015; Stange, 2012). In this model, the student trades off between spending a time taking courses across different majors in order to make a more informed major choice, with concentrating her course-taking to graduate as quickly as possible.

Finally, the model provides a microfoundation for the cost of switching majors. Focusing her course-taking in one major hastens graduation, but the student builds large switching costs because it will take relatively more courses to graduate in other majors. Diversifying her course-taking increases information about major ability, and keeps switching costs from inhibiting future choices.

Formally, the student makes an amount of progress, $\mathbb{c}_{imt}$, in major m at semester t. She receives flow payoffs ν_m from each major she makes progress in. When she accumulates at least 100% progress in a major, she graduates in that major and receives a graduation payoff, $Grad_{imt}$.

2.4.1 College Flow Payoffs and Graduation Payoffs

The flow payoff depends on the grade the student receives g_{imt} and amount of progress the student makes in that major, $\mathbb{c}_{imt}$.[6] The flow payoff is characterized with the following comparative statics:

$$\frac{\partial \nu_{imt}}{\partial g_{imt}} > 0, \quad \frac{\partial \nu_{imt}}{\partial \mathbb{c}_{imt}} > 0, \text{ and } \frac{\partial^2 \nu_{imt}}{\partial \mathbb{c}_{imt}^2} < 0 \tag{2.1}$$

(2.1) assumes students receive a net benefit from earned grades. The model explicitly assumes the grade yields a benefit net of effort costs.

The major-specific flow payoff is concave in progress $\mathbb{c}_{imt}$. For lower levels of $\mathbb{c}_{imt}$, the

[6]There is no constraint on how much progress she can make across all majors. Unmodeled effort costs and logistical limitations can constrain how much progress the student can make in a semester. Other work examining course choice holds the number of attempted credits each period constant (Hendricks and Leukina, 2015).

student's flow payoff increases independent of grades, for reasons such as taking courses with peers or course experiences with major m. However, her marginal utility from making additional progress decreases for higher levels of progress. This can be because she receives lower marginal benefits from non-grade flow payoffs or her total effort costs convexly increase. The concavity assumption for ν_{imt} gives the student incentive to diversify her course-taking in each time period and also holds the student back from graduating early.

Upon completing major m's course requirements, the student receives graduation payoff $Grad_{imt}(\overline{g_{imt}})$, where $\overline{g_{imt}}$ is the student's grade point average (GPA) in major m. The graduation payoff represents a combination of labor market returns to graduation and other non-pecuniary benefits such as leisure and marriage (Wiswall and Zafar, 2013). Ties of completing multiple majors are randomly broken among majors she simultaneously finishes in.

$$\frac{\partial Grad_{imt}}{\partial \overline{g_{imt}}} > 0 \tag{2.2}$$

$Grad_{imt}$ does not depend on the student's progress in other majors. When the student is close to finishing one major, she may make progress in other majors for the flow payoffs. Graduation payoffs indirectly depend on her unknown innate abilities because she forms expectations of future grades.

2.4.2 Ability Learning Framework

As the student makes progress across majors, she receives major-specific grades. The student believes grades are a function of her innate major-specific abilities. From the student's beliefs, she forms expectations of her future grades, $\mathbb{E}[g_{imt}] = b_{imt}$. The difference between her realized and expected grades is a random expectation error, composed of the idiosyncratic errors in grades and errors in beliefs.

Following Bayesian (DeGroot, 1979) models on worker learning (Altonji and Pierret, 2001), the random expectation error has two components. The first is that there is random noise around the grade, and the second is the difference between the student's actual and believed ability. I assume that the student has rational expectations of her beliefs, and there is individual variation in students' ability beliefs.

The student uses the difference between her major-specific GPAs and initial beliefs in major abilities to update her beliefs $b_{ti} = (b_{i1t}, b_{i2t}, \cdots, b_{iMt})$. The student weights the difference between initial beliefs and major-specific GPAs with her cumulative progress across majors. Initial beliefs b_i^o are based on time-invariant characteristics, X_i.

$$b_{im}^o = \phi_m X_i + \epsilon_{im}^b \tag{2.3}$$

ϵ_{im}^b is an idiosyncratic error of her initial beliefs the econometrician does not observe. The student also believes that her abilities are correlated across majors (Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2014b). Then the student updates her beliefs in majors $k \neq m$ from a grade in major m.

Since the student does not know her major abilities, she has an information incentive to experiment with different majors. She can also exploit the correlation between majors and make progress in a major she believes is highly correlated with others. Expectation errors cause revisions of the student's beliefs and change her choices over time.

2.4.3 Model Results for Empirical Testing

The course-taking model's main features are how students take courses to learn about their major abilities and complete majors' course requirements. Then the student's state variables are her initial beliefs, GPAs, and cumulative progress: $S_{it} = (b_i^o, \overline{g_{it}}, \overline{\mathbb{c}_{it}})$. These state variables calculate her future expected flow payoffs, graduation payoffs, and beliefs in her major-specific abilities. The student continues making choices only if she does not drop out and has not finished any majors. Then her value function is expressed as:

$$\begin{aligned} \mathrm{V}_t(S_{it}) \quad &= \max\{0, \max_{\mathbb{c}_{it}} \Big\{ \overbrace{U_{it}}^{\textit{Flow Payoff}} + \beta \Big[\overbrace{\sum_{m=1}^{M} \mathbb{1}\{K = m\} Grad_{im\tilde{t}}) +}^{\textit{Finish Any Major}} \\ &\underbrace{(1 - \mathbb{1}\{K = m\}) \cdot \mathbb{E}\big[\mathrm{V}(S_{it+1}) | S_{it}\big]}_{\textit{Finish No Majors}} \Big] \Big\}\} \end{aligned} \tag{2.4}$$

where $U_{it} = \sum_{m=1}^{M} \nu_{imt}(g_{imt}, \mathbb{c}_{imt}) \mathbb{1}\{\mathbb{c}_{imt} > 0\}$.

If the student has no uncertainty about her abilities, her objective becomes a static problem and her strategy would be one of two extremes. The first is graduating as soon as possible in order to attain the graduation payoff. The second is to stay in college as long as possible to attain the positive flow payoffs nu_m before graduating in one major. Then in the static problem, there is little incentive to diversify course-taking in order to insure against changes in ability beliefs, and switching costs do not play any role.

Under the learning framework, the student updates her beliefs as she earns grades. The higher her earned grades are than her previous belief, her ability belief increases. These beliefs directly impact her valuation of expected payoffs. As she receives more grades, she attributes more of the expectation error to be idiosyncratic shocks. She grows more certain of her ability beliefs, and the effect of subsequently earned grades dampen over time.

$$\frac{\partial b_{imt+1}}{\partial \overline{g_{imt}}} > 0, \quad \frac{\partial^2 b_{imt+1}}{\partial \overline{g_{imt}} \partial \overline{\mathbb{c}_{imt}}} < 0 \tag{2.5}$$

The cross-major comparative statics capture how beliefs are correlated across majors. If majors are not correlated, then these cross-major comparative statics are zero. Likely these cross-major derivatives are weakly positive - the student believes all majors are positively correlated.

The comparative statics on beliefs carry over to her choices. There are optimal policy functions for how much progress she makes in each major $e^*_{it} = (e^*_{i1t}, e^*_{i2t}, \cdots, e^*_{iMt})$. Grades enter the student's choice through her beliefs b_{it}. Since expected future flow and graduation payoffs increase with beliefs, comparative statics with respect to her beliefs in own major m are:

$$\begin{aligned} \frac{\partial e^*_{imt}}{\partial b_{imt}} &> 0 \Rightarrow \frac{\partial e^*_{imt}}{\partial g_{imt}} > 0 \\ \frac{\partial e^*_{imt}}{\partial b_{ikt}} &\leq 0 \text{ for } k \neq m \end{aligned} \tag{2.6}$$

Comparative statics in (2.6) come from the learning comparative statics in (2.5) – as the student's major-specific GPA increases, she increases her expectation of future flow and graduation payoffs in that major and makes more progress in that major. The second comparative static is how getting higher beliefs in one major decreases the incentive to make progress in other majors.

Results on own-major do not directly translate to comparative statics on cross-major grades, $\frac{\partial e^*_{imt}}{\partial g_{ikt}}$, because of the correlated framework. If the student believes majors m and k are highly correlated, then the student would upwardly revise her beliefs in both abilities after receiving a higher grade in major k. Even then it is unclear whether the student will make more progress in major m when she receives a higher grade in major k, and this depends on the relative payoffs in majors m and k.

As the student makes more progress in one major and is closer to finishing it, she has a greater incentive to continue. As the student makes progress across multiple majors, her incentive to continue making progress across majors decreases because she only receives graduation payoffs from graduating in one major[7] and graduating earlier. Another incentive to persist is that the switching cost may be large: the student does not want to restart taking courses in another major.

$$\begin{aligned} \frac{\partial e^*_{imt}}{\partial \mathbb{C}_{imt}} &> 0 \\ \frac{\partial e^*_{imt}}{\partial e_{ikt}} &< 0, \text{ for } k \neq m \end{aligned} \tag{2.7}$$

However, it is not clear from the model if the student should accelerate how much progress she makes as she gets closer to finishing, $\frac{\partial^2 e^*_{imt}}{\partial \mathbb{C}_{imt}{}^2} > 0$. This is because once the student finishes

[7]If the student can double-major, she has an incentive to make progress in multiple majors. There is still the trade-off of graduation time.

a major's course requirements, then she graduates college and can no longer learn about her major-specific abilities.

From these comparative statics, I expect students to make more progress in a major as their major-specific GPA increases. It is unclear whether or not students will similarly respond to grades in other majors. Their response depends on how majors are correlated and their relative payoffs.

2.5 Testing the Model's Comparative Statics

With students constantly making choices depending on their previous academic experiences and grades, it is difficult to rely on naturally exogenous variation in grades across all the majors. On the course-taking front, while students are required to take a series of courses to satisfy general distributional requirements, they can take them in any semester. The course-taking model provides a framework to test how students change their course-taking in response to different grades, and how students' cumulative progress anchors them into different majors.

This estimation strategy is not designed to causally estimate the magnitude and direction of the comparative statics. I focus on the relationship between course grades and persistence in taking courses needed to graduate in different majors. To account for how students may be exploring other majors, and the incentive to graduate in other majors, I control for observed grade histories and progress across majors. I use a rich set of student covariates to account for students' individual preferences to graduate in majors: ACT scores, SAT scores, high school GPA, pre-college interest.

2.5.1 Sample Selection

The ideal sample is a population of students who have an identical amount of information about their major-specific abilities before entering CALS and do not enroll in CAS with the intention to transfer to other schools. Therefore, I drop students with more than 24 transfer credits, transfer from CAS to another PHEC school within their first two semesters.

Students with transfer credits have information incomparable to those from PHEC courses. Advanced Placement (AP) exam scores and grades from other institutions are drawn from different distributions.[8] Students with significant transfer credits from two-year institutions likely took courses with the intention of transferring, and students may take AP courses to improve their admission likelihood. Many schools at PHEC such as the Business, Education, and Public Policy schools recruit their entire student does from CAS.[9] Then some students enroll in CAS with the intention of internally transferring to another PHEC school. I use leaving within

[8] While PHEC accepts credits from outside institutions and AP exams, their grades do not carry over and transfer students start with empty cumulative grade point averages.

[9] The Business School started directly admitting high school students around 2007, but still recruits a substantial portion of its students as internal transfers from CAS.

two semesters as a conservative proxy for students' intentions to leave CAS before enrolling (approximately 1% of the final sample).

Table 2.3 show descriptive statistics for the sample selection steps. I drop students who enter CAS with more than 24 transfer or AP credits and leave CAS after two semesters. Compared to the final sample, students entering with more than 24 transfer credits and AP credits are more likely to be male, and Asian. Dropped students also score around ten percentiles highers in the ACT and SAT Math and Reading sections.

2.5.2 Estimating Equations

I estimate logit reqressions[10] on whether students make any progress in different majors.[11] The outcome of interest is whether the student makes progress in major m as a function of the ability signals the student receives, and her cumulative progress in all majors.

I use the difference between the students' major GPAs and a separately estimated initial belief of major-specific abilities as grade signals. From the Bayesian learning framework, the relevant information to the student is not the grade itself, but the grade relative to the student's initial belief.[12] Using cumulative GPAs, $\overline{g_{imt}}$, naturally captures how students' beliefs should become more resilient as they receive more grades. I rely on estimates of students' initial beliefs from regressing polynomials of student's time-invariant characteristics on first semester GPAs. I define the difference between the student's cumulative GPA and the initial belief, as "Grade Shocks."

$$GS_{imt} = \overline{g_{imt}} - \widehat{b_{im}^{o}} \tag{2.8}$$

where $b_{im}^{o} = \phi_m X_i + \epsilon_{im}^{b}$. I regress whether students make any progress in a major on "Grade Shocks" m GS_{imt} and progress $\overline{\mathbb{c}_{imt}}$ across all majors. The outcome variable y_{imt} indicates whether or not the student makes any progress in major m in semester t.

$$\begin{aligned} y_{imt} = \quad & \text{logit}\Big(\sum_{k=1}^{M} \big[\alpha_{mk} GS_{ikt} + \gamma_{mk}\overline{e_{ikt}} + \delta_{mk} GS_{ikt} \times \overline{e_{ikt}}\big] + \\ & \omega_m \,\text{Withdraws}_{imt} + \pi_m X_i + \tau_t\Big) \end{aligned} \tag{2.9}$$

Where X_i are the students' baseline characteristics: reported gender, race, ACT and SAT Math and Reading percentile scores, 2012 median zipcode earnings, prior interests in major m. τ_t are semester fixed effects, which represent how students may be financially or otherwise con-

[10] I also estimate ordinary least squares linear probability regressions in Table B.3, showing similar results.

[11] Table B.1 shows the most common choice across students' semesters is to make no progress in that major. Aside from Humanities (36%), in more than 50% of observed student-semesters students make no progress in at least one of these majors.

[12] An alternative specification is to fully interact X_i with the cumulative GPAs. Since initial beliefs are time-invariant, it is preferable to interpret estimates as relative to prior beliefs instead of how the influence of cumulative grades vary over prior beliefs.

strained from enrolling in additional semesters.

α_{mk} and γ_{mk} test the implications of the model. As students' "Grade Shocks" increase, they upwardly revises beliefs in major abilities. As students make more progress, this increases the incentive to finish that major. The comparative statics (2.6) and (2.7) predict $\alpha_{mm} \geq 0$ and $\gamma_{mm} \geq 0$.

The opportunity cost of transitioning course-taking into another major should decrease the more progress students have made on that major. This predicts estimates of progress across majors to be negative, $\gamma_{mk} \leq 0$, but is ambiguous on whether receiving positive signals in one major increases or decreases the change of making progress in other majors. Ambiguity comes from correlation in major abilities, and relative payoffs from course-taking and graduating across majors.

The model has an ambiguous prediction for the estimate on the interaction term. The interaction between "Grade Shocks" and cumulative progress captures how students' beliefs become resilient to subsequent "Grade Shocks." Following the Bayesian learning framework, students with more positive "Grade Shocks" in a major will make smaller revisions to their major ability belief, suggesting $\delta_{mm} < 0$ and $\delta_{mk} < 0$. However, the interaction term also captures how students who have made more progress respond when they receive higher "Grade Shocks," with predicts a positive relationship, suggesting $\delta_{mm} > 0$ and $\delta_{mk} > 0$

I also measure students' ability signals with cumulative course withdraws in that major each semester. Withdraws are recorded in the administrative data if the student withdraws from the course after the third week of the semester, and are calculated as a failing grade. I interpret withdraws as negative signals of ability, and predict $\omega_m < 0$.

2.6 Results and Discussion

Table 2.4 underlines the importance of including information across majors. Without conditioning on "Grade Shocks" from other majors, students with higher "Grade Shocks" in the Humanities and Social Sciences are less likely to make any progress in those majors. A correlation between Humanities and Social Sciences abilities will downward bias these estimates, as students with high Humanities and Social Sciences abilities likely have high abilities in other majors.

I report parameters estimates in Table 2.5 and focus on the direction of the parameters. Again, these estimates are not meant to be causal and test the dynamic course-taking model's comparative statics. Students take courses to make progress in majors' course requirements as a function of their measured "Grade Shocks" to their initial beliefs, and previous progress across majors.

2.6.1 Estimates on Earned "Grade Shocks"

Parameter estimates on students' "Grade Shocks" over time are consistent with the model. Students with higher own–major "Grade Shocks" are more likely to make progress in those majors, and previously withdrawing from courses in that major is negatively related with making progress.

Whether students make progress in one major is negatively related to having higher "Grade Shocks" in other majors. This is consistent with students' payoffs from taking courses and graduating in majors to be positively related to grades. Students are drawn away from majors because they believe they have higher abilities and hence higher payoffs in other majors. In some cases, the cross-major estimates positive and suggests correlated major abilities: students with higher Psychology "Grade Shocks" are more likely to make progress in the Math & Physical Sciences, Life & Physical Sciences, and Economics majors. Students with higher Psychology "Grade Shocks" not only believe they have higher abilities in other majors, but that higher abilities are rewarded more in these other majors.

2.6.2 Estimates on Cumulative Progress

Once students have already completed some of the course requirements for a major, they have an incentive to continue. Table 2.5 shows students who have made more progress and completed more of the course requirements for a major are more likely to persist. The incentive to diversify course-taking across majors seems to pale against against graduating quickly.

When students switch majors, they transition their course-taking across majors. Using students' progress in other majors provides one of the first looks into the microfoundations for these switching costs. Parameter estimates suggest the more progress students have made in one major, the less likely they are to make progress in another major. This is consistent with Table 2.2, where students make more progress in majors they previously declared. The increasing switching costs highlights the importance of first year courses, which can set students on the path to graduating in a major, independent of major ability belief.

Several positive estimates suggest double-majoring or simultaneously making progress across majors. Students with more progress in Psychology are more likely to make any progress in the Life & Earth Sciences, and vice versa. This symmetric relationship suggests Psychology and Life & Earth Sciences send each other students, and the switching cost is relatively small. There are also asymmetric relationships between majors, indicating on average the Math & Physical Sciences and Economics send students to Economics and Psychology, respectively.

2.6.3 Estimates on "Grade Shock" and Progress Interactions

As students make more progress and earn more grades, the model predicts their major ability beliefs become more resilient. Negative estimates on the interaction between "Grade Shocks" and progress support this prediction. With fewer revisions to their major ability beliefs, students are less likely to make any progress in majors than if they made more progress.

Positive estimates on the interaction of Life & Earth Sciences "Grade Shocks" and progress shows students who perform better and have made more progress in these majors make progress in other majors. This is reduced-form evidence that the Life & Earth Sciences is sending high ability students to other majors.

2.6.4 Estimates on Time Invariant Characteristics

Controlling for "Grade Shocks" and progress across majors accounts for previous experience and beliefs in major abilities. Estimates on student characteristics capture how certain students made progress in their first semester.[13] Looking at student characteristics, I find that Female students are more likely to make progress in the Humanities and Social Sciences. This is consistent with previous work on gender gaps in major choice (Brown and Corcoran, 1997), though the gaps are not substantially significant. Compared to White students, Black students are more likely to make any progress in the Math & Physical Sciences, Life & Earth Sciences, Social Sciences, and Psychology. Asian students are less likely to make any progress in the Math & Physical Sciences and more likely to make any progress in the Humanities. Students with higher Reading percentile ACT or SAT scores are more likely to continue in the Humanities, while students with higher Math percentile scores are more likely to continue in the Math & Physical Sciences, Life & Earth Sciences, and Economics.

Students' prior interests also play a role in how students progress through majors. These interests can come from a host of reasons, including family (Anelli and Peri, 2015; Zafar, 2012) and high school experiences (Darolia and Koedel, 2016). These prior interests generally play a positive role in students' course-taking decisions. Students interested in the Social Sciences prior to enrollment are less likely to make any progress in the Social Sciences major.[14] The negative estimate suggests that after controlling for students' grade and course histories, students may be learning they have a lesser preference for the Social Sciences.[15] If these interests are inherently misinformed, then the negative estimate for the Social Sciences suggests students are learning over time.

[13]Suppose Female students were always more likely to make progress than Males in each semester, then controlling for cumulative progress account for this time invariant pattern.

[14]The average marginal effect from Table B.2 is negative 0.8 percentage points, which is relatively smaller than other factors to making any progress in the Social Sciences.

[15]Figure B.1 shows the influence of prior interest in the Social Sciences is positive until students' sixth semester, even without controlling for "Grade Shocks" and progress. This suggests students are learning about their Social Sciences ability independently from "Grade Shocks." Overall, it demonstrates the waning influence of time invariant characteristics over time.

2.6.5 Discussion & Conclusion

With ongoing public and policy attention on influencing major choice, this paper is one of the first to use transcript data to study the correspondence between course-taking and major choice. The paper finds descriptive evidence that students are lacking information about their major ability, which calls into question previous works that students are not making fully informed major choices. Excluding the Arts and Economics, students have completed between 10% to 20% of the course requirements needed to graduate in all majors before declaring their major. While the administrative data does not include earnings outcomes, the literature has found substantial determinants beyond earnings and major ability beliefs (Wiswall and Zafar, 2013).

After declaring their major, students continue to take a substantial amount of courses in other majors, completing on average 10% to 25% of the course requirements needed to graduate in the Social Sciences, Humanities, and Psychology. With students taking courses relevant to majors other than their declare ones, major declaration is likely not suitable to measuring students' likelihood of switching majors.

A model of course-taking resolves issues of using major declarations to understand transcript data. In the model, students have an incentive to diversify their course-taking in order to insure against low ability belief revisions, and to prevent substantial switching costs from limiting their future choices. I find the data is consistent with the model's comparative statics: students with higher "Grade Shocks" are more likely to take courses and complete course requirements in a major, and students are less likely to take courses in a major if they have completed the course requirements for other majors.

Rather than taking courses across multiple majors throughout college to keep the opportunity cost of switching majors low, students quickly concentrate their course-taking. This is consistent with students wanting to quickly graduate and that switching costs become salient early on. The empirical results stress the complexities that course-taking policies must address: assigned course-taking may creates inertia to persist in a major independent of major ability beliefs, and the waning influence of grades on influencing course-taking. On the other hand, early course-taking can create a counterfactual inertia to continue taking the courses needed to graduate in a major. This is relevant to constructing future policies, as first year courses can set students onto the path to graduating in a major.

Taken altogether, the transcript evidence shows students face conflicting incentives to graduate earlier and learn about major abilities, while facing the risk of learning they have low ability in a major they have already taken many courses in. These issues arise from the higher education institution's design goal to let students explore different majors. Yet, courses, particularly those required to graduate for different majors, may play a substantial role in setting students on the path to graduating in a major independent of ability.

Table 2.1: Major Groupings

- **Math & Physical Sciences**
  Computer Science
  Actuarial Math
  Honors Math
  Pure Math
  Mathematical Sciences
  Statistics
  Astronomy
  Physics

- **Life & Earth Sciences**
  Biochemistry
  Biology
  Biomolecular Science
  Biophysics
  Cellular Molecular Biology
  Chemistry
  Ecology and Evolutionary Biology
  General Biology
  Microbiology
  Neuroscience
  Plant Biology
  Earth Sciences
  Earth and Environmental Sciences

- **Social Sciences**
  Anthropology
  Evolutionary Anthropology
  Information Science
  Political Science
  Politics, Philosophy, and Economics
  Sociology

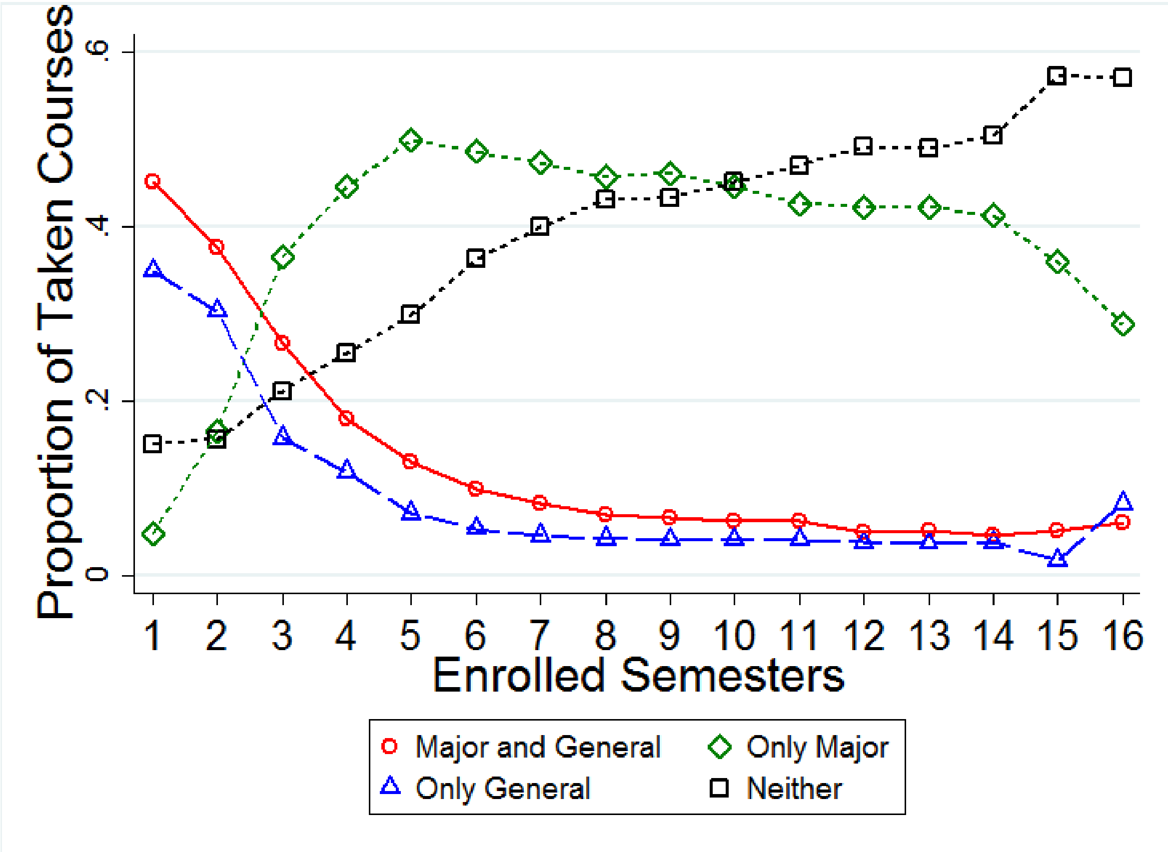
- **Economics**
  Economics

- **Psychology**
  Psychology
  Organizational Behavior
  Biopsychology, Cognition, and Neuroscience

- **Humanities**
  Afroamerican and African Studies
  American Culture
  Art and Ideas
  Communications
  Comparative Literature
  General English
  Honors English
  Judaic Studies
  Middle Eastern Studies
  Asian Studies
  International Studies
  Latin Studies
  Linguistics
  Philosophy
  Social Theory and Practices
  Women's Studies
  Classical Architecture
  Classical Civilizations
  Classical Languages
  Ancient Civilizations and Biblical Studies
  Ancient Greek
  History
  Art History
  French
  German
  Modern Greek
  Italian
  Polish
  Russian
  Spanish

- **Arts**
  Creative Writing
  Drama
  Music
  Screen Art Culture

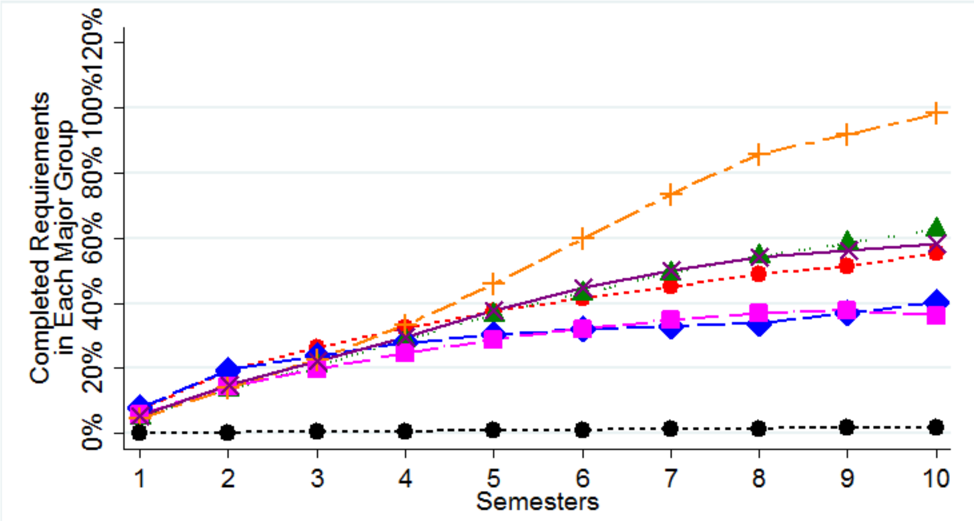
Figure 2.1: Overlap in Majors' and Distributional Course Requirements



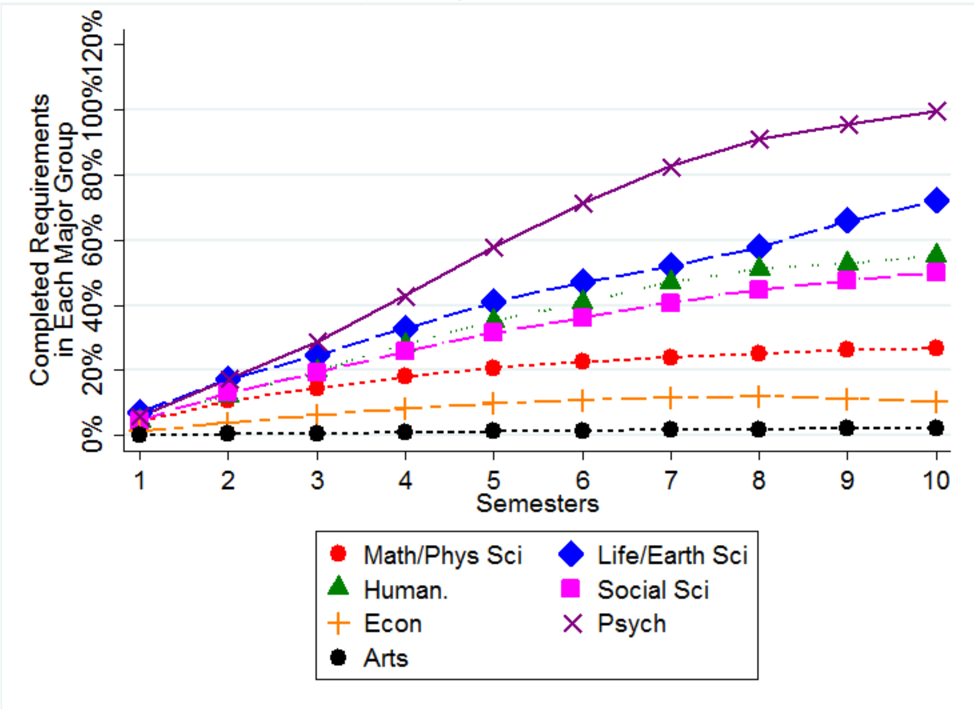
NOTES – This figure uses all taken courses at PHEC, in addition to courses taken in other schools at PHEC. Transfer credits are not known. Courses that students withdraw from are not used in this figure.

Figure 2.2: Cumulative Annual Progress for Economics and Psychology Graduates

Economics Graduates:

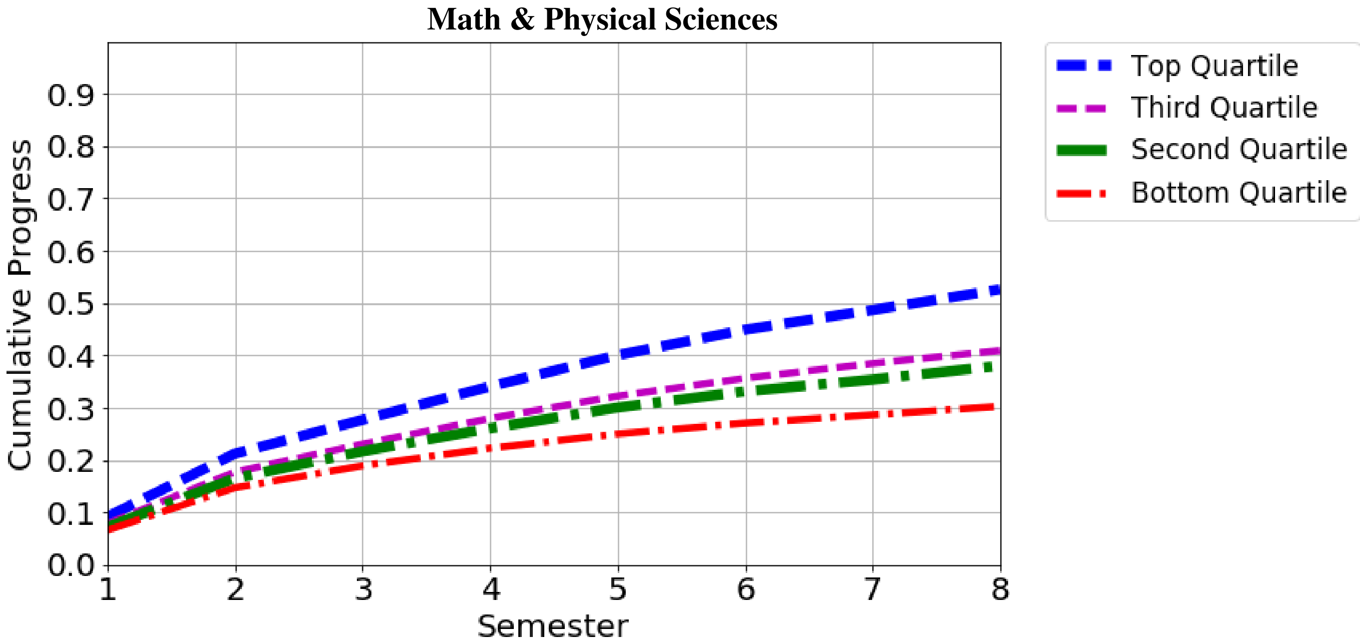


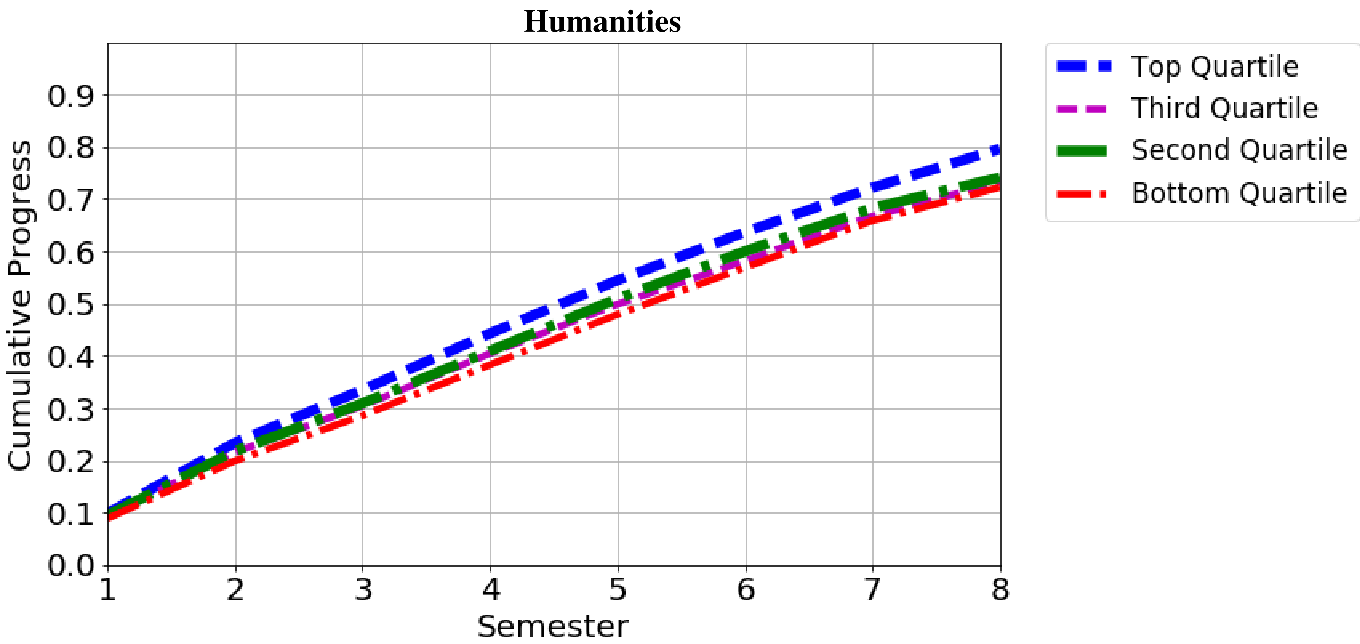
Psychology Graduates:



NOTES – Average semester cumulative progress is shown here . Students are included if they are inferred as ever completing the major based on coursework. The first half of the Summer Semester is included with the Spring Semester and second half included with the Fall Semester. Cumulative progress is carried forward for students who leave PHEC.

Figure 2.3: Cumulative Progress over Own-Major First Year GPA Quartiles





NOTES– Quartiles are calculated based on courses defined as related to the major. These courses are beyond the required classes to complete the major. Students who did not take these courses in their first semester are excluded from these statistics. Cumulative progress is carried forward for students who leave PHEC.

Table 2.2: Amount of Progress Students Have Made in Majors' Course Requirements Before and After Declaring in a Major

Declared Major	One Semester Before Declaring: Cumulative Progress Across Majors							
	Math & Physical Sci.	Life & Earth Sci.	Humanities	Social Science	Economics	Psychology	Arts	Average Semester of Declaration
Math & Physical Sciences	0.40	0.37	0.19	0.19	0.07	0.14	0.01	4.60
Life & Earth Sciences	0.19	0.48	0.19	0.23	0.04	0.22	0.00	4.70
Humanities	0.11	0.14	0.30	0.18	0.06	0.17	0.01	4.28
Social Science	0.13	0.16	0.24	0.30	0.07	0.19	0.01	4.44
Economics	0.26	0.23	0.25	0.22	0.26	0.28	0.01	4.94
Psychology	0.13	0.18	0.25	0.21	0.07	0.35	0.01	4.56
Arts	0.08	0.10	0.24	0.13	0.03	0.12	0.16	4.29

Declared Major	Four Semesters After Declaring: Change in Cumulative Progress Across Majors						
	Math & Physical Sci.	Life & Earth Sci.	Humanities	Social Science	Economics	Psychology	Arts
Math & Physical Sciences	0.45	0.11	0.16	0.14	0.05	0.11	0.01
Life & Earth Sciences	0.09	0.36	0.18	0.22	0.02	0.28	0.01
Humanities	0.04	0.07	0.44	0.15	0.04	0.14	0.02
Social Science	0.05	0.07	0.29	0.41	0.04	0.16	0.01
Economics	0.10	0.05	0.21	0.11	0.41	0.20	0.01
Psychology	0.04	0.10	0.24	0.13	0.03	0.36	0.01
Arts	0.03	0.03	0.26	0.06	0.02	0.07	0.33

NOTES – The table treats making progress and declaring a major simultaneously occurring within a semester. Students are tracked to make progress until they leave PHEC.

Table 2.3: Sample Selection

	More than 24 Transfer Credits	More than 24 AP Credits	Missing ACT/SAT	Leave within Two Semesters	Final Sample
Female	0.443	0.426	0.667	0.548	0.590
	(0.497)	(0.495)	(0.479)	(0.501)	(0.492)
Black	0.0314	0.00277	0.0909	0.0548	0.0634
	(0.175)	(0.0526)	(0.292)	(0.229)	(0.244)
Asian	0.159	0.293	0.303	0.219	0.126
	(0.365)	(0.455)	(0.467)	(0.417)	(0.332)
Hisp	0.0386	0.0185	0.0303	0.0274	0.0543
	(0.193)	(0.135)	(0.174)	(0.164)	(0.227)
Reading Percentile	84.62	93.42	—	83.37	83.19
	(13.44)	(6.689)	(—)	(15.76)	(14.80)
Math Percentile	90.58	95.28	5.697	85.76	85.24
	(10.37)	(5.427)	(22.78)	(16.48)	(14.05)
Transfer Credits	40.79	1.994	11.82	7.685	2.592
	(15.63)	(3.873)	(20.59)	(6.205)	(4.163)
AP Credits	8.977	29.69	2.939	7.904	5.881
	(9.323)	(6.128)	(6.164)	(7.707)	(6.527)
Other Credits	0.764	0.382	0.727	0.123	0.696
	(2.481)	(2.057)	(2.908)	(0.600)	(2.241)
N	6678	2161	33	72	36647

NOTES – Standard deviations are reported in parentheses. Sample selection steps start from the left and go to the right. ACT and SAT scores are aggregated using percentile scores, with the average being taken when both are available.

Table 2.4: Own Shock and Progress Estimates for Different Groups of Controlling Variables, from OLS Regressions

	Whether the Student Makes Any Progress in Major						
	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Including Demographic Characteristics, Semester Fixed Effects:, and Own Progress:							
Own Shock	0.011***	0.001	−0.008**	−0.004	0.040***	0.013***	0.031**
	[0.003]	[0.002]	[0.003]	[0.003]	[0.003]	[0.002]	[0.010]
R^2	0.200	0.367	0.149	0.161	0.305	0.157	0.231
Include Grade Shocks from All Other Majors:							
Own Shock	0.028***	0.005	0.044***	0.037***	0.082***	0.066***	0.042**
	[0.005]	[0.005]	[0.006]	[0.005]	[0.005]	[0.005]	[0.016]
R^2	0.205	0.380	0.155	0.177	0.307	0.166	0.232
N	266012	254085	254168	265793	268643	264814	269426

NOTES– $*** = p < 0.01$; $** = p < 0.05$, $* = p < 0.10$
Standard errors are clustered on the student level and reported in brackets. The sample in each regression is restricted to student-semester observations where students' cumulative progress is less than 100%.

Table 2.5: Extensive Margin of Making Progress in A Major
Logit Parameter Estimates

	Whether the Student Makes Any Progress in Major						
	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Grade Shocks Across Majors and Withdraws							
Own Shock	0.186***	0.019	0.220***	0.193***	0.671***	0.301***	0.545*
	[0.028]	[0.026]	[0.027]	[0.024]	[0.049]	[0.023]	[0.261]
Own Withdraws	–0.132***	–0.110***	–0.268***	–0.272***	–0.224***	–0.279***	–0.265***
	[0.016]	[0.017]	[0.014]	[0.014]	[0.023]	[0.014]	[0.045]
Math & Phys Sci. Shock		0.065*	–0.036	–0.042	0.001	–0.037	0.066
		[0.029]	[0.024]	[0.025]	[0.035]	[0.023]	[0.068]
Life & Earth Sci. Shock	–0.085***		–0.107***	–0.114***	–0.213***	–0.192***	0.03
	[0.024]		[0.021]	[0.021]	[0.031]	[0.021]	[0.059]
Humanities Shock	0.008	0.083**		–0.024	–0.074*	–0.043	0.101
	[0.029]	[0.030]		[0.025]	[0.037]	[0.024]	[0.071]
Social Sci. Shock	–0.055*	–0.070*	0.080***		–0.005	0.014	0.022
	[0.027]	[0.028]	[0.023]		[0.035]	[0.023]	[0.066]
Economics Shock	–0.095**	–0.091*	–0.090**	–0.087**		0.084**	–0.158
	[0.034]	[0.038]	[0.030]	[0.033]		[0.030]	[0.084]
Psych. Shock	0.119***	0.179***	–0.015	–0.016	0.174***		–0.074
	[0.025]	[0.026]	[0.022]	[0.022]	[0.035]		[0.062]
Arts Shock	–0.088	–0.125	–0.097	–0.077	0.041	–0.101	
	[0.124]	[0.131]	[0.097]	[0.102]	[0.181]	[0.099]	
Cumulative Progress Across Majors							
Own Cumul. Prog	5.417***	6.177***	3.716***	4.021***	9.229***	3.966***	9.245***

Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
	[0.053]	[0.042]	[0.026]	[0.027]	[0.075]	[0.024]	[0.143]
Math & Phys Sci. Cumul Prog.		−1.263∗∗∗	−0.473∗∗∗	−0.024	0.517∗∗∗	−0.333∗∗∗	0.06
		[0.052]	[0.032]	[0.038]	[0.059]	[0.031]	[0.106]
Life & Earth Sci. Cumul. Prog	−0.083∗∗		−0.341∗∗∗	0.746∗∗∗	−0.338∗∗∗	0.586∗∗∗	−0.097
	[0.029]		[0.021]	[0.023]	[0.042]	[0.021]	[0.070]
Humanities Cumul. Prog	−0.791∗∗∗	−0.912∗∗∗		−0.571∗∗∗	−0.787∗∗∗	−0.373∗∗∗	−0.336∗∗∗
	[0.030]	[0.034]		[0.024]	[0.044]	[0.022]	[0.075]
Social Sci. Cumul Prog.	−0.226∗∗∗	−0.609∗∗∗	0.063∗		−0.286∗∗∗	0.397∗∗∗	−0.291∗∗∗
	[0.031]	[0.037]	[0.025]		[0.046]	[0.023]	[0.078]
Economics Cumul. Prog.	−0.139∗∗	−0.950∗∗∗	−0.380∗∗∗	−0.707∗∗∗		0.542∗∗∗	−0.165
	[0.045]	[0.059]	[0.034]	[0.042]		[0.033]	[0.123]
Psych. Cumul. Prog.	−0.292∗∗∗	0.732∗∗∗	−0.349∗∗∗	−0.093∗∗∗	−0.348∗∗∗		0.219∗∗
	[0.030]	[0.035]	[0.023]	[0.023]	[0.044]		[0.079]
Arts Cumul. Prog.	−1.355∗∗∗	−1.542∗∗∗	−0.680∗∗∗	−1.375∗∗∗	−1.354∗∗∗	−0.920∗∗∗	
	[0.103]	[0.114]	[0.063]	[0.079]	[0.144]	[0.067]	
Grade Shock and Cumulative Progress Interactions Across Majors							
Own Shock × Cumul. Prog	−0.151	0.149∗	−0.424∗∗∗	−0.360∗∗∗	−1.657∗∗∗	−0.549∗∗∗	−0.546
	[0.086]	[0.061]	[0.066]	[0.058]	[0.167]	[0.057]	[0.813]
Math & Phys Sci. Shock × Cumul. Prog		−0.181∗	−0.273∗∗∗	−0.119	−0.271∗∗	−0.12	−0.394∗
		[0.092]	[0.060]	[0.077]	[0.104]	[0.061]	[0.189]
Life & Earth Sci. Shock × Cumul. Prog	−0.133∗∗		0.149∗∗∗	0.182∗∗∗	−0.002	0.205∗∗∗	0.184
	[0.048]		[0.039]	[0.039]	[0.076]	[0.039]	[0.132]
Humanities Shock × Cumul. Prog.	0.03	−0.041		0.033	0.262∗	−0.035	−0.760∗∗∗
	[0.074]	[0.080]		[0.059]	[0.106]	[0.054]	[0.169]

Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Social Sci. Shock × Cumul. Prog	–0.028	–0.016	–0.213***		–0.101	–0.152**	–0.126
	[0.074]	[0.083]	[0.058]		[0.104]	[0.056]	[0.175]
Economics Shock × Cumul. Prog	0.231*	–0.02	–0.025	0.124		–0.316***	0.446
	[0.103]	[0.138]	[0.081]	[0.104]		[0.081]	[0.280]
Psych. Shock × Cumul. Prog.	–0.157*	–0.230**	–0.139**	–0.059	–0.260**		0.133
	[0.068]	[0.072]	[0.053]	[0.055]	[0.098]		[0.166]
Arts Shock × Cumul. Prog.	0.628	1.19	0.353	0.711	0.692	1.093*	
	[0.591]	[0.631]	[0.376]	[0.445]	[0.862]	[0.425]	
Time Invariant Characteristics							
Female	–0.190***	–0.051***	0.038***	0.032***	–0.229***	–0.023**	–0.222***
	[0.011]	[0.013]	[0.009]	[0.010]	[0.015]	[0.009]	[0.028]
Black	0.110***	0.089***	–0.038*	0.149***	–0.024	0.137***	0.061
	[0.024]	[0.026]	[0.018]	[0.020]	[0.035]	[0.018]	[0.062]
Asian	–0.035*	–0.035	–0.070***	0.056***	–0.049*	–0.009	0.058
	[0.016]	[0.020]	[0.013]	[0.014]	[0.023]	[0.013]	[0.040]
Hisp	0.062**	0.027	0.020	0.114***	0.032	0.035*	0.027
	[0.023]	[0.027]	[0.019]	[0.020]	[0.032]	[0.018]	[0.060]
Reading Percentile	–0.001**	–0.003***	0.002***	–0.001*	–0.002***	–0.003***	0.002*
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.001]
Math Percentile	0.005***	0.002**	–0.005***	–0.004***	0.009***	–0.001**	–0.001
	[0.001]	[0.001]	[0.000]	[0.000]	[0.001]	[0.000]	[0.001]
2012 Median Zipcode Income	0.014	0.000	0.022**	0.014	0.003	0.001	0.010
	[0.009]	[0.010]	[0.007]	[0.008]	[0.012]	[0.007]	[0.022]
Prior Interest	0.092***	0.375***	0.119***	–0.040***	0.116***	0.031*	0.608***
	[0.015]	[0.014]	[0.009]	[0.012]	[0.026]	[0.013]	[0.049]

Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Constant	−5.026***	−0.889	−2.258	−3.628***	−0.642	−3.249***	−6.257***
	[1.500]	[1.207]	[1.447]	[0.851]	[1.169]	[0.934]	[0.864]
Psuedo-R^2	0.199	0.340	0.124	0.147	0.334	0.134	0.296
log–likelihood	-119908.204	-113487.508	-152868.818	-153986.882	-76041.288	-158020.193	-30365.683
N Clusters	36297	36290	36292	36297	36297	36297	36297
N	265999	254083	254164	265786	268589	264808	269258

NOTES –

$* * * = p < 0.01; \;\; * * * = p < 0.05, \;\; * = p < 0.10$

The calculated shocks to these majors are the difference of estimated initial ability beliefs from cumulative major GPAs. Initial priors are estimated using a flexible polynomial of pre-collegiate time-invariant characteristics. A complete list of these covariates can be found in the Appendix. Standard errors are clustered on the student level and reported in brackets. Entry term and semester fixed effects are not reported. In addition to the sample restriction in Table 2.3, each regression is restricted to student-semester observations where students' cumulative progress is less than 100%.

Chapter 3

Math for All? Regression Discontinuity in Signals of Preparation for College Quantitative Coursework

with William J. Gehring, University of Michigan, Department of Psychology

3.1 Introduction

3.1.1 Addressing the Needs of Under–Prepared Students

A key challenge for American higher education is to accommodate the different levels of pre-college preparation among high school graduates as they enter college. Some students arrive at college lacking the background needed for introductory coursework, whereas others arrive already possessing college credit in the form of Advanced Placement and other courses. A critical task for colleges is to provide opportunities for less-prepared students to rise to the level necessary to compete with their better-prepared peers. If colleges fail to meet this need, they risk widening the existing inequities in post-college outcomes for students whose backgrounds have failed to prepare them adequately for college. Chen and Weko (2009) find that academically less-prepared students are less likely to graduate college in a Science, Technology, Engineering, or Math (STEM) field. In this paper, we find recommendations to take Pre-Calculus and Calculus are not adequate tools to narrow these gaps in selective four-year colleges.

Despite the body of evidence on remedial mathematics at two-year colleges (Bettinger and Long, 2005, 2009; Boatman and Long, 2010; Horn et al., 2009; Martorell and McFarlin, 2011; Melguizo et al., 2015; Scott-Clayton and Rodriguez, 2015; Scott-Clayton et al., 2014), there is a dearth of evidence on the academic trajectories of less-prepared students at selective four-year colleges. The need for such evidence is strong because of the labor market benefits of enrolling at a four-year institution (Hoekstra, 2009; Jaegar and Page, 1996; Kane and Rouse, 1995).

Preparatory coursework can help less-prepared students achieve outcomes similar to their more prepared peers. These courses likely play a role in how higher college quality improves the academic outcomes of less-prepared students (Arcidiacono and Lovenheim, 2015; Dillon and Smith, 2017). In addition, there is evidence that less-prepared students divert from more difficult majors into easier majors to achieve the same grade-point outcomes as their better-prepared peers (Arcidiacono et al., 2012a). Where course difficulty comes from students' preparation levels, preparatory courses that provide the foundation for advanced courses are important. Hence, understanding the mechanisms driving student course and major choices has important policy implications.

3.1.2 Studying Institutional Recommendations to take Pre-Calculus and Calculus

Students can learn about their capacity to succeed in different kinds of courses from their actual course experiences. In particular, students learn about their ability to succeed in college math and other STEM classes once they have taken a math class. There is evidence that students' course and major choices are motivated by their anticipated grades (Bar et al., 2009; Main and Ost, 2014). Thus, researchers must be careful in taking self-selection into account for drawing proper inferences. We avoid these selection biases by exploiting discontinuities in an institutional recommendation policy. These recommendations occur before the first day of class and are among the first interactions students have with their institution, and are an excellent opportunity for students to learn about their capacity for success in a course without spending the effort (or financial resources) to take courses.

In this paper, we focus on four mutually exclusive recommendations to take Pre-Calculus or Calculus at a public flagship four-year institution, which we anonymize as the Flagship Academic College, FAC. We exploit discontinuities in the administrative formula used to assign recommendations in a regression discontinuity design to estimate the causal impacts of these Pre-Calculus and Calculus recommendations on student course-taking over time. Each recommendation (Pre-Calculus or Calculus) can come with two degrees of firmness (Tentative or Definite), yielding four mutually exclusive recommendations: Pre-Calculus/Definite Recommendation; Pre-Calculus/Tentative Recommendation, Calculus/Tentative Recommendation, and Calculus/Definite Recommendation.

We use a relatively new non-parametric procedure (Calonico et al., 2014) to estimate the causal Intent-to-Treat (ITT) impacts of receiving different recommendations. We leverage our transcript data to decompose student course-taking in different subjects and find that students who marginally receive the "Tentative Recommendation to take Calculus" rather than the "Tentative Recommendation to take Pre-Calculus" are 15 percentage points more likely to take Calculus by their second semester. By the eighth semester, these differences disappear, indicating the marginal students given the "Tentative Recommendation to take Pre-Calculus" have caught up with their peers.

Among the least-prepared students (those receiving one of the two Pre-Calculus recommendations), those who marginally receive the "Tentative Recommendation to take Pre-Calculus" are 17 percentage points more likely to ever take Calculus. They are also less likely to ever take a course that fulfills the quantitative graduation requirement but does not count toward a major at FAC. Across all recommendations, we find limited impacts on course-taking trajectories in taking courses that require Calculus, and more general subjects: Economics, Statistics, Biology, or Chemistry. This suggests students do not take advantage of their induced Calculus experience in future courses and inducing students to take Calculus is insufficient to narrowing STEM gaps.

3.1.3 Expanding the Literature: Recommendations Instead of Placements

Our work focuses on detailed college course-taking, similar to how Melguizo et al. (2015) looked at persistence in college course sequences. While these outcomes are important to college graduation and progress to degree, it is perhaps less so at our elite public institution, which has among the highest persistence and graduation rates among public institutions.[1] Consequently, our study focuses on more fine-grained outcomes that can reveal changes in course-taking trajectories. Another difference in our context is that the recommendation we use has several levels, allowing us to study the effects of recommendations on students of different preparation levels. Because we study a four-year institution, our analysis also follows students into longer and more diverse academic tracks than Melguizo et al. (2015). The rich options available to students in a four-year university allow us to study how the recommendations influence students' subsequent decisions to take quantitative courses and courses having quantitative prerequisites.

We view the placement recommendation as an informational treatment, similar to the approach taken in the growing literature using regression discontinuity designs based on Advanced Placement (AP) exam scores Papay et al. (2011); Smith et al. (2017). This literature finds that, independent of granted AP credits, students with almost identical abilities are more likely to major in fields related to the exams in which they scored higher. Recommendations are an informational treatment, giving students information about their likelihood of success in quantitative coursework. Students can also interpret this recommendation as an indication of how they stand relative to their peers. While students with AP credits at FAC also receive Pre-Calculus or Calculus recommendations, they likely respond very differently from less-prepared students without AP credits. For example, their AP credits can reduce the need for prerequisite courses prior to STEM classes. Students taking AP exams are likely not under-prepared, and are not in the population of interest in our paper.

To reiterate, our paper studies students' reactions to "Definite" and "Tentative" recommendations to take Pre-Calculus and Calculus. Unlike the developmental (remedial) literature where students are placed into a developmental course, we do not evaluate the impacts of taking Pre-Calculus or Calculus. While Scott-Clayton and Rodriguez (2015) consider the

[1] Revealing actual statistics also reveals FAC's identity.

informational impacts of being placed into remedial courses, our context is quite different. Remedial courses do not count towards graduation credit, and taking them does not get students closer to graduating. FAC does not offer remedial courses; thus the Pre-Calculus and Calculus courses both count toward graduation credit and satisfy the distributional quantitative graduation requirement.[2] In addition, students are free to ignore this recommendation and take other courses that satisfy the same distributional course requirement.

Our paper also differs from that prior work in that there is no natural control group, as all students receive recommendations. Thus, we estimate the relative effects of receiving different recommendations. Students are not told how recommendations are calculated, and they do not have an incentive to adjust their recommendation because the recommendation is not binding.

3.2 Institutional Context and Data Description

3.2.1 Pre–Calculus and Calculus at our Institution of Study

In our paper, we use administrative transcript data from students at a highly selective public four-year institution, Flagship Academic College (FAC). FAC houses many colleges and schools; the two largest are its College of Arts and Sciences (CAS) and Engineering College. The College of Arts and Sciences has the largest student body (approximately 60% of total undergraduate enrollment) and provides students with the most diverse set of courses. CAS students can take courses outside of the CAS, and college majors such as Business, Education, and Public Policy are not housed in CAS.

The transcript data records each course students take while at FAC, including transfer credits the student earned prior to enrollment at FAC, and transfer credits the student transferred in while attending FAC. Courses are not recorded if a student withdraws from a course within the first three weeks of the semester. The transcript separately records attempted and earned credits, as well as the letter grade received from each course.[3]

All CAS students are required to take one preapproved course to satisfy a quantitative distributional graduation requirement. Some of these preapproved courses, such as Linear Algebra, Introduction to Microeconomics, and Introductory Statistics, also satisfy other majors' course requirements and are clear stepping stones to more advanced courses. Other courses that satisfy the distributional requirement do not satisfy any major's course requirements, and do not pave the way to more advanced courses. For example, there are courses focusing on the history of DNA, cryptology, and looking for extraterrestrial life. We refer to these courses as "Non-Major Quantitative Courses" and investigate them as a separate outcome. AP credits do not satisfy

[2] Because this course counts towards overall graduation credit, the "diversion" effect in Scott-Clayton and Rodriguez (2015) is not relevant in our context. Although we find substitution patterns between Pre-Calculus and other quantitative courses, students will satisfy the quantitative course requirement regardless of their choice.

[3] Less than 3% of all student-courses taken at CAS result in withdrawals, which are recorded if the student withdraws later than three weeks after the start of the semester.

the quantitative graduation requirement.

Pre-Calculus and Calculus also satisfy this quantitative distributional requirement. Despite being called “Pre-Calculus,” the Mathematics Department at CAS designs the Pre-Calculus curriculum to be self-contained and not a stepping stone to Calculus.

All students who enter FAC, regardless of previous mathematics background or transfer credits, receive one four mutually exclusive recommendations:

1. Definite Recommendation to take Pre-Calculus (DP)
2. Tentative Recommendation to take Pre-Calculus (TP)
3. Tentative Recommendation to take Calculus (TC)
4. Definite Recommendation to take Calculus (DC)

These recommendations are based on a Math Index calculated from each student’s high school grade point average, math SAT or ACT score, and math placement exam score. Students take the math placement exam before starting any coursework.[4] Three explicit cutoffs in the Math Index (at 1.5, 2.0, and 2.25) determine students’ recommendations. Further details can be found in the Appendix. Here, we refer to these four recommendations as DP, TP, TC, and DC.

At a pre-orientation event[5] the summer or winter before they start coursework, students meet with academic advisors who provide them with this recommendation. This is the only time students observe these recommendations: they are not presented to the student after taking the math placement exam, and cannot be otherwise found afterwards. FAC wants students to take these recommendations seriously. If students express reluctance, academic advisors try to convince students to follow the recommendations. For example, if the student wanted to take Calculus instead of following a Pre-Calculus recommendation, the advisor could show the student a Calculus textbook and challenge the student to comprehend the first few chapters. Ultimately, students can still choose to ignore their advisors’ recommendations.

3.2.2 Sample Selection of Students

We use administrative data from FAC to calculate this index and recreate the recommendations. Although all students receive these recommendations, we are concerned with examining the largest population of students for whom these recommendations have the same amount of

[4] In 2003, the math placement exam moved from a pencil-and-paper exam students took at the pre-orientation event before enrolling to an online exam students could take any time before the pre-orientation event. We could not find differences in recommendations following this change, though our data only starts in 2002. The Math Index is actually based on a regression calculated a number of years ago that predicts the grade a student would receive in calculus (on a four-point scale) if her or she were to take it without additional preparation.

[5] Students have a variety of dates to choose from during the summer, and require a separate orientation fee and accommodation expenses. Students who do not attend the event during the summer are required to attend a functionally identical event a few days before the start of the semester.

relevance. Therefore, we limit our sample to students who enter FAC through CAS from 2002 to 2008 as freshmen and did not have any transfer credit for Pre-Calculus or Calculus. We track these students for six years after they enter CAS.

We show the sample selection in Table 3.1. Focusing our analysis on students who initially enroll at FAC through CAS, we drop students who initially enrolled through the Engineering College, which comprises proportionally fewer female students and students with higher SAT and ACT math scores. The other dropped students are transfers from outside institutions, who enter with lower math achievement measures, and have a higher proportion of Black, Asian, and Hispanic students. Students who are dropped for having Advanced Placement Calculus credit for Calculus I are more likely to be male and Asian and to have higher SAT and ACT scores.

3.2.3 Descriptive Statistics of Recommendations and Quantitative Course–Taking

FAC does not record the continuous (running or forcing) variable used to compute the recommendation. Therefore, we computed the running variable and recommendation using the formula and cutoffs provided by the registrar. Figure 3.1 shows that we make few errors in recreating them. Despite extensive effort, we were unable to determine why we were unable to perfectly recreate students' recommendations.[6] We retain students in our sample regardless of whether we correctly recreated their recommendations, giving us a fuzzy regression discontinuity design. We consequently estimate Intent-to-Treat (ITT) effects.

An obvious result of the placement recommendation is that students might alter the ways they fulfill the quantitative course requirement. Although these recommendations are for Pre-Calculus and Calculus, which satisfies the quantitative course requirement, students can also choose from a list of classes that meet the quantitative requirement but do not satisfy any major's course requirements.[7] We refer to this group of courses as "Non-Major Quantitative Courses."

Table 3.3 gives a preliminary look at the outcome variables indicating whether or not students ever take different quantitative courses by their second, eighth, and twelfth semesters. We consider three groups of quantitative courses that count towards FAC's quantitative distributional requirement:

1. Pre–Calculus
2. Calculus

[6] The CAS administration proposed two reasons for this inconsistency: The first is that the administrative data is overwritten as new data appears - students who retake the ACT or SAT have their data rewritten in the administrative records. The second is inconsistent reporting of high-school GPA and other scores across different data tables.

[7] There are also courses that satisfy the quantitative course requirement which also satisfy different majors' course requirements. We take these courses into account in examining students' course-taking in different subjects.

3. “Non-Major Quantitative Courses”

Table 3.3 shows almost no differences between the proportions of students who take these courses between their eighth and twelfth semesters. With this in mind, we look only at outcomes by students’ second and eighth semesters. These semesters are calculated as subsequent semesters after initial enrollment; very few students skip semesters.

The descriptive evidence supports the hypothesis that the recommendations affected students’ course choices. Table 3.3 shows that the proportion of Definitely Recommended to Take Pre-Calculus (DP) students who take Pre-Calculus increases from 40% in the second semester to 44% in the eighth semester. Students Tentatively Recommended to Take Pre-Calculus (TP) behave similarly, with the proportion taking Pre-Calculus increasing from 36% to 38%.

Looking at the proportion of students who take Calculus, we see that TC and DC students take Calculus in much higher proportions, with 54% of TC and 43% of DC students taking Calculus by their eighth semester. Taking Calculus is likely to be more popular among TC students than DC students in part because DC students are better prepared to take more advanced quantitative courses.

Taking Non-Major Quantitative Courses is less popular than taking Pre-Calculus or Calculus. DP students take Non-Major Quantitative Courses in higher proportions than all other students, around 21% by the end of their eighth semesters. At the other extreme, only 12% of DC students take Non-Major Quantitative Courses. From their second to eighth semesters, the proportion of students who take Non-Major Quantitative Courses increases around ten to fifteen percentage points across all recommendations.

3.3 Estimation Strategy

Comparing these descriptive statistics most likely results in biased inferences about the causal impact of recommendations. DP students are likely very different from TP students in unobserved ways: previous experiences with quantitative coursework, desires to pursue STEM majors, study habits, and quantitative ability. These unobserved differences likely correlate with our outcomes of interest: taking Pre-Calculus, taking Calculus, graduating in STEM majors, and course-taking trajectories.

To estimate the causal impact of these recommendations, it is helpful to think about the ideal experiment where students are randomly given Definite or Tentative recommendations to take Pre-Calculus or Calculus recommendations or none at all. Using the calculated Math Index, we separate students into four levels of quantitative preparation. For the least prepared students, we would randomly assign students to receive a Definite Recommendation to take Pre-Calculus (DP) or none at all. This assignment process would continue, until for the most prepared students, we randomly assign students to receive a Definite Recommendation to take

Calculus (DC) or none at all.[8] With students randomly assigned recommendations suitable for their level of quantitative preparation or none at all, we could estimate the causal impact of each of the four recommendations.

3.3.1 Interpretation of Receiving Different Recommendations

Our context differs from the ideal experiment, because all students are given a recommendation. Thus, we can only estimate the relative effects of different recommendations. If we simply compare two groups with different recommendations, and one group shows a greater measured outcome than the other, it is not possible to know whether one recommendation increased the outcome or the other recommendation decreased it, or both.[9] Observing a difference between those students confounds two distinct reactions to the recommendation. We interpret relative effects of these different recommendations as students' reactions to signals of their quantitative preparation.

In contrast to most educational interventions, we draw attention to the somewhat peculiar properties of placement recommendations and evaluating a context where all students are treated. Unlike contexts where the policy recommendations are whether to treat or not, our setting compares the relative impacts of different recommendations. In our regression discontinuity context, the most obvious policy question is how the cutoff should be placed.

Robinson (2011) discusses how the cutoff should be placed with respect to finding null impacts. If the intent of Pre-Calculus and Calculus recommendations are to close gaps in STEM course-taking or major choices, then an estimated null impact means there is either no discernable impact or the cutoff has been placed to equalize outcomes. It is likely these recommendations are not intended to close gaps, and we focus on unintended consequences of recommending students to taking Pre-Calculus or Calculus.

We present a framework based on Robinson (2011) for understanding the cost and benefits of setting the cutoff. Suppose the university is interested in equalizing outcomes between students with less (L) and more (H) preparation on some outcome Y. Suppose $Y_L < Y_H$, and the university assigns some treatment (T) based on students' preparation to equalize outcomes. If the university perfectly set the cutoff such that only L students are assigned the treatment, then we would find no differences using the RD design: $Y_L + T = Y_H$. If the cutoff is too low, then not enough L students are assigned the treatment. We would find the treatment has a positive effect, though outcomes are not equalized.

[8] It is likely the highest achieving students would ignore any recommendations to take Pre-Calculus. Policy makers would likely not be interested in the impacts of Definitely Recommending the least prepared students to take Calculus. Both scenarios have obvious ethical shortcomings.

[9] This consideration is relevant in our present scenario. An anecdotal report from an academic advisor said that students getting the TP recommendation, below the recommendation to take Calculus, often argue for a higher placement, whereas students getting the TC recommendation often argue for a lower one. It is also not clear why students argue for different recommendations, as recommendations can be ignored without consequence.

If the cutoff is too high, then H students are assigned. If T_H is positive, such as offering studying strategies, then the university can afford to set the cutoff too high. We would find a positive effect of the treatment at the margin: $Y_H + T_H > Y_H$. Yet the RD design will not capture the policy margin of interest, between L and H students. L and H outcomes may not be equalized (this depends on the relative magnitudes of T_L and T_H).

We now depart from Robinson (2011)'s approach and consider if T_H is negative,[10] such as recommending students to take a developmental course. Taking this developmental course could slow H students from graduating on time. If the cutoff were set too high, we would then find a negative effect of the treatment at the cutoff: $Y_H + T_H < Y_H$, even though the treatment benefits less prepared students, $T_L > 0$. The university could incorrectly either the treatment does not work, or (following Robinson (2011)) the cutoff is too low. The case where $T_H < 0$ is important for the university to consider. If it sets the cutoff too high, it pays a cost of decreasing the outcomes of H students.

3.3.2 Regression Discontinuity Validity and Approach

We use the recommendations based on explicit cutoffs in a regression discontinuity (RD) design. RD estimates causal impacts by comparing students who just above or below along a continuous quantitative dimension, in our case, those who receive different recommendations. RD is used in various fields, including Economics, Education, and Psychology (Cook, 2008), as a compelling method to estimate causal impacts.

We study four mutually exclusive recommendations, and estimate relative impacts of marginally receiving one recommendation compared to another. Since recommendations are based on cutoffs in the calculated Math Index, we can only make three comparisons. We would estimate the three equations below, each on students in a narrow interval around the discontinuity.

$$Y_{it} = {}_0 + {}_1 MathIndex_i + \alpha_1 \mathbb{1}\{\text{Tentative Recommendation to Take Pre-Calculus}_i\} \quad (3.1)$$

$$Y_{it} = {}_2 + {}_3 MathIndex_i + \alpha_1 \mathbb{1}\{\text{Tentative Recommendation to Take Calculus}_i\} \quad (3.2)$$

$$Y_{it} = {}_4 + {}_5 MathIndex_i + \alpha_1 \mathbb{1}\{\text{Definite Recommendation to Take Calculus}_i\} \quad (3.3)$$

If the regression discontinuity design is valid, then α_1 estimates the impact of the Tentative Recommendation to take Pre-Calculus relative to the Definite Recommendation to take Pre-Calculus (TP – DP), α_2 estimates the impact of Tentative Recommendation to take Calculus relative to the Tentative Recommendation to take Pre-Calculus (TC – TP), and α_3 estimates the impact of the Definite Recommendation to take Calculus relative to the Tentative Recommendation to take Calculus (DC - TC). These estimates are for students at the cutoffs, who are marginally given a higher recommendation or signal of quantitative preparation.

[10]This case is not considered in Robinson (2011), as it examines for English Learner Reclassification, where the treatment effect is assumed to be positive.

(3.1), (3.2), and (3.3) are estimated using a subsample of students with Math Index values in a narrow interval around the cutoff values. We use a fuzzy regression discontinuity design, where α_1, α_2, and α_3 estimate the Intent-to-Treat (ITT) impacts on students who marginally receive these recommendations (Lee and Lemieux, 2009).

RD requires two tests that provide evidence in support of this causal interpretation. The first test is that students are as good as randomly assigned in narrow intervals around the cutoffs for different recommendations. One reason this may not happen is if students tried to manipulate their calculated Math Index. This occurs in other contexts when the formula is publicly known (Lee, 2001). We use a McCrary density test (McCrary, 2008) which tests the null hypothesis that the Math Index running variable has a continuous density around the cutoffs. We run separate tests for each cutoff and entering cohort year. Our results appear in Table 3.2;[11] they show that the density in the Math Index is generally smooth through the cutoffs. We find differences that approach statistical significance, at the 10% level, in TP - TC (comparing that Math Index density for students Tentatively Recommended to take Pre-Calculus against those who are Tentatively Recommended to take Calculus). We keep this in mind moving forward, but describe below why this should not be interpreted as evidence manipulation.

Students are not aware of how their Math Index value is calculated, and there is no incentive to manipulate their recommendations. Recommendations are shared between the student and advisor at the pre-orientation event, and are not binding. Neither the student nor the advisor ever sees the calculated Math Index. While students can re-take the math placement exam, we only calculate the Math Index value using students' first math placement exam score. Very few students retake the math placement exam: only 22 students (less than 0.1%) in our entire sample. We argue students are not trying to manipulate their scores for desirable recommendations.

The second test is whether there is balance in student control variables at the cutoff. If there are discrete jumps in student control variables, then any difference in outcomes at the cutoff in the running variable can be attributed to both the jump in student control variables and the treatment. We plot out student characteristics over the index in Figures 3.4 3.5, 3.3, and 3.2. Figures C.1 and C.2 plot across other student characteristics. We find no visual evidence of any jumps at the cutoff.

We rigorously test for differences in student control variables, using the same non-parametric regression discontinuity estimator from Calonico et al. (2014). Table 3.4 shows students who are marginally given the Tentative Recommendation to take Pre-Calculus, rather than the Definite Recommendation to take Pre-Calculus are less likely to come from earlier cohorts. Students marginally given the Tentative Recommendation to take Calculus, rather than the Tenta-

[11] We show the McCrary test for the unrounded math index values. Although the formula used to calculate the math index rounds to the nearest hundredths place, the cutoffs are 1.50, 2.00, and 2.25 and sensitive to rounding. Our preliminary analyses showed that the McCrary test failed using the rounded math index. This does not affect our arguments against manipulation, and the rounding creates a data artifact uncorrelated with potential confounds (Barreca et al., 2016).

tive Recommendation to take Pre-Calculus are more likely to be Asian and score higher on the SAT Math section. Students marginally given the Definite Recommendation to take Calculus, rather than the Tentative Recommendation to take Pre-Calculus are less likely to be White, more likely to not report a race, more likely to report multiple races, score higher on the ACT Math and Science sections, and have lower high school GPAs.

What can cause these discontinuous jumps in student control variables? Recall the calculated Math Index is calculated using high school grade point average, math SAT or ACT score, and math placement exam score. Then discontinuous jumps in enrollment year and reported race is based on their correlation with these scores. Discontinuous jumps in ACT and SAT section scores, and high school grade point averages likely come from the availability of these scores. Section C.1 discusses the availability of these scores.

We now estimate the impact of receiving different recommendations on students. We examine three different cutoffs, comparing:

1. Definite Recommendation to take Pre-Calculus against Tentative Recommendation to take Pre-Calculus (TP – DP);
2. Tentative Recommendation to take Pre-Calculus against Tentative Recommendation to take Calculus (TC – TP); and
3. Tentative Recommendation to take Calculus against Definite Recommendation take Calculus (DC - TC).

In applying the parametric analysis shown in (3.1), (3.2), and (3.3), we found that estimating TP – DP, TC – TP, and DC - TC was highly sensitive to the interval or bandwidth selection. As the bandwidth decreases, the estimate becomes more susceptible to random noise: it becomes under powered. As the bandwidth increases, the estimate becomes more precise because the variance decreases. However, a larger bandwidth runs the risk of including students who are not on the margin of being assigned the treatment, thereby invalidating the RD design. We also found that the estimates were sensitive to the different functional forms of (3.1), (3.2), and (3.3): polynomials of the running variable and interactions with the cut-off.

There is no definitive method to determine the best interval or parametric form.[12] To address this issue while allowing us to "let the data speak for itself", we use non-parametric regression and a kernel-weighting function of the distance of observations from the cutoff. However, this method also relies on bandwidth selection. Unfortunately, the bandwidth selection required to approximate the kernel-weighting function is often too large and introduces the bias of involving students who are more likely to be different in unobserved ways. Too large a bandwidth

[12] In Figures C.4 and C.3, we plot the running variable over whether students take Pre-Calculus or Calculus. Linear regressions along the entire interval imply differences we do not see in the non-parametric specifications.

can lead us to incorrectly reject the null hypothesis the parameter is not different from zero. As Calonico et al. (2014) observe, “This is a well-known problem in the nonparametric curve estimation literature” (pg. 2301). We use a method to calculate robust non-parametric confidence intervals presented in Calonico et al. (2014) that corrects for this bias using variation from the larger bandwidth. A separate “pilot bandwidth” is used to calculate a separate bias estimate that is subtracted from the original estimate. In all tables showing estimates, we present the estimate, P-value, bandwidth (not the “pilot bandwidth”), control mean (on the left-hand side of the cutoff), and sample size from this bias-corrected process.

3.4 Non-Parametric Regression Discontinuity Results

In our non-parametric RD estimates, we estimate the effect of receiving the marginally higher recommendation. For example, along the TP – DP margin, we estimate the Intent-to-Treat (ITT) estimate of the marginal effect of receiving the Tentative Recommendation to take Pre-Calculus (TP) compared to the Definite Recommendation to take Pre-Calculus (DP) for students at the margin. Table 3.3 showed little variation between students’ eighth and twelfth semesters, and we show ITT estimates on students’ outcomes by the end of their second and eighth semesters.

3.4.1 ITT Impacts on Cumulative Credits and Graduation

One of the main outcomes of interest in the developmental course literature is students’ cumulative credits over time (Horn et al., 2009; Martorell and McFarlin, 2011; Scott-Clayton et al., 2014). Cumulative attempted credits roughly measures how students make progress toward completing graduation requirements. We show non-parametric RD estimates on cumulative credits on Table 3.5. We measure students’ cumulative credits by the second and eighth semesters, as well as the difference in cumulative credits between students’ eighth and second semesters. For students who drop out or leave the institution before these semesters, we carry their cumulative credits forward rather than give them zeros or drop them from the sample.

We find that along the margin of receiving the Definite Recommendation to take Pre-Calculus (DP) and Tentative Recommendation to take Pre-Calculus (TP), TP students take more credits in their second semester. The estimated effect sizes are neither statistically nor substantively significant. Along the TC – TP and DC - TC margins, we also find that marginal students given higher recommendations to take Calculus neither statistically nor substantially increase attempted cumulative credits.

Yet we find recommendations affect course withdrawals, with estimates shown on Table 3.5. Students at the TP – DP margin withdrew from half a course more by their eighth semester when they were marginally given the Tentative Recommendation to take Pre-Calculus. The 65% increase (0.568 on a control mean of 0.853) difference is statistically and substantially significant. This result suggests recommendations cause students to overreach and enroll in courses they are not prepared for. One reason we may see this during students' eighth semesters is that these courses are not necessarily related to Pre-Calculus or Calculus.

Table 3.5 also shows RD estimates on whether the student graduates by their eighth semester. We find that students who marginally received the Tentative Recommendation to take Pre-Calculus (TP) are four percentage points less likely to graduate by their eighth semester compared to students who marginally received the Definite Recommendation to take Pre-Calculus (DP). We find a 4.7% decrease (4.5 on 96.5) that statistically significant at the 10% level. We find this estimate surprising, as students given marginally higher signals of their quantitative preparation are more likely to drop out. This result differs from the literature on remedial course-taking. In formulating an interpretation, one could argue that the higher recommendation could lead to a riskier course-taking strategy, with some number of students attempting courses that do not match their preparation level. Estimates on graduation for TC relative to TP and DC relative to TC are statistically and substantively insignificant.

3.4.2 ITT Impacts on Quantitative Course–Taking

Across all recommendations, we find statistically and substantially insignificant impacts on whether students take Pre-Calculus. However, we find impacts on taking Calculus. We see students who marginally receive the Tentative Recommendation to take Pre-Calculus compared to the Definite Recommendation to take Pre-Calculus are more likely to ever take Calculus by their second and eighth semesters. These differences are substantial and statistically significant: marginal TP students are 16 percentage points more likely to take Calculus in their first or second semester, on a control mean of 22 percent. This difference persists to the eighth semester: TP students are 18 percentage points more likely to ever take Calculus by their eighth semester, on a control mean of 27 percent. This shows recommendations affect students' early decisions to take Calculus, and few students on the TP – DP margin revisit this decision (as they might if the experience in Pre-Calculus were sufficiently positive to encourage them to take Calculus). To summarize this finding bluntly, FAC's placement recommendation, rather than equalizing Calculus outcomes for the least quantitatively prepared students, causes them to be less likely to take Calculus.

We find a short-lived impact on Calculus taking for the marginal TP and TC students. Students marginally given the Tentative Recommendation to take Calculus (TC) are 15 percentage points more likely to take Calculus by their second semester, on a control mean of 38 percent. By their eighth semesters, this difference is statistically insignificant but substantially significant: 7 percentage points more likely on 47 percent. The TP students who receive a lower signal about their quantitative preparation are initially discouraged from taking Calculus, but catch up with their TC student peers who received a higher signal about their quantitative preparation.

Yet is taking Calculus rewarding for these students? We find suggestive evidence these recommendations have a negative influence on students' Calculus grades. Estimates on students' first earned grades on 3.7 show marginal TP and TC students score 0.3 points lower (approximately from a "B+" to "B"). While these estimates are not causal because they are conditional on students taking Calculus, it suggests the calculus course itself can be dissuading students from continuing into quantitative majors.

3.4.3 ITT Impacts on Taking Non-Major Quantitative Courses

We next look at other reactions to these recommendations. Upon getting a recommendation, students can elect to take a course other than Pre-Calculus or Calculus to satisfy the quantitative requirement.

The top portion of Table 3.6 shows whether students ever take a Non-Major Quantitative Course. The marginal TP students are 2 percentage points less likely to take Non-Major Quantitative Courses by their second semester compared to the marginal DP students. This is statistically significant at the 5% level. The 2 percentage point difference is substantial as the control mean is 2 percent, and the difference persists into the eighth semester, where TP students are 12 percentage points less likely to take these courses, on a control mean of 26 percent. The level growth in this effect between semesters 2 and 8 is significant, suggesting that the disparity in taking Non-Major Quantitative Courses grows over time. This shows that among the less quantitatively prepared students, the signals of lower quantitative preparation conveyed by the recommendation causes students to take more Non-Major Quantitative Courses. The result on ever taking Calculus is of a similar magnitude, suggesting students are substituting Calculus with Non-Major Quantitative Courses.

3.4.4 ITT Impacts on Courses Requiring Pre–Calculus and Calculus

We have seen that these recommendations have substantial impacts on whether students ever take Calculus and Non-Major Quantitative Courses. We then move to measure impacts on taking courses that require Pre-Calculus or Calculus. Courses that require Calculus include most Physics courses, intermediate and advanced Economics, intermediate Statistics, Chemistry, and Math courses.

ITT estimates on the top of Table 3.8 show that along all recommendation margins, TP – DP, TC – TP, and DC - TC, we do not find statistically significant differences. However, the direction of these estimates are consistent with students responding to higher signals about their quantitative preparation for college courses. Students who marginally receive higher recommendations are more likely to take courses that require Pre-Calculus or Calculus.

3.4.5 ITT Impacts on Cumulative Credits in STEM Related Subjects

Another outcome of interest is whether students use these recommendations to draw inferences about their ability or capacity to continue in more quantitative fields. Such a finding would be particularly informative to policies that wish to increase student participation in certain quantitative or Science, Technology, Engineering, and Math (STEM) fields, if students are not being properly encouraged about their quantitative preparation.

We look at the cumulative credits students have taken in Chemistry, Biology, Mathematics, Statistics, and Economics in Tables 3.8 and 3.9. Courses are usually three or four credits each, and Calculus is a prerequisite for majors in these fields. Recommendations to take Calculus can open up these fields for students. We find short-run differences in cumulative attempted Mathematics, Statistics, and Economics credits, but overall find substantially and statistically insignificant results across recommendations.

Table 3.8 shows estimates on cumulative attempted Chemistry and Biology credits by students' second and eighth semesters. Across all recommendations, we find statistically insignificant differences. Along the TC – TP margin, we find positive differences for students' cumulative Chemistry credits, suggesting the marginal Tentative Recommendation to take Calculus has some positive influence on Chemistry course-taking.

In Table 3.9, we find statistically significant results along the TP – DP margin for cumulative second semester Math and Economics credits, and along the DC - TC margin for cumulative second semester Statistics credits. We find that giving Tentative Recommendations to take Pre-Calculus increases cumulative attempted second semester Math and Economics credits by around one credit (0.7 on control means of 2.4 and 0.629). These Math and Economics credit differences are significant at the 10% and 5% levels, respectively. Compared to students who marginally receive Definite Recommendations to take Pre-Calculus, TP students have attempted 100% more Economics credits by their second semester. However, these differences are not statistically significant by students' eighth semesters.

Along the DC - TC margin, students who marginally receive Definite Recommendations to take Calculus take 0.5 more Statistics credits (on a control mean of 0.7) by their second semesters. Like previous results, this difference also decreases by students' eighth semesters. These results on Math, Economics, and Statistics suggests that at best, recommendations can have short-lived impacts on coursework in quantitative fields.

3.4.6 ITT Impacts on STEM Major Completion

Finally, we find that recommendations to take Pre-Calculus or Calculus have no statistically discernable impact on a student's graduating major. Table 3.10 shows that recommendations have no statistically significant impact on whether students graduate in Biology, Economics, Physics, Mathematics, or Statistics. The estimates suggest a substantive difference in Economics: DP students are 5 percentage points more likely to graduate in Economics (control mean of 3 percentage points) relative to TP students. Results on graduating majors are consistent with our findings on course-taking in different subjects.

3.5 Discussion and Conclusion

In this paper, we use administrative student transcripts and admissions data from a four year public institution to estimate the causal impacts of receiving Tentative or Definite Recommendations to take Pre-Calculus or Calculus on taking quantitative college courses over time.

Our paper makes its largest contribution to the growing literature of how students respond to information about their ability and preparation to major in certain fields (Papay et al., 2011; Smith et al., 2017). We examine how student course-taking responds to institutional recommendations across all students, rather than responses conditional on taking AP exams. We also focus on a different population of students, who did not receive Advance Placement Exam credit and who are less prepared for college work.

We interpret recommendations as signals about students' quantitative preparation, and find favorable signals about students' quantitative preparation cause them to be more likely to take Calculus. Marginal students who receive the Tentative Recommendation to take Pre-Calculus are 85% more likely to take Calculus in their first or second and eighth semesters than those who receive the Definite Recommendation. We find neither substantially nor statistically significant differences in whether students ever take Pre-Calculus.

We also find that recommendations affect whether less-prepared students take Non-Major Quantitative Courses, which satisfy a quantitative distributional requirement but do not fulfill course requirements in any major. Students who receive the Tentative Recommendation to take Pre-Calculus are 100% less likely to ever take Non-Major Quantitative Courses than if they receive the Definite Recommendation to take Pre-Calculus. Although this difference decreases over time, it remains statistically and substantially large, showing early information sets students on a course-taking trajectory away from the courses required for quantitative majors. Taking Non-Major Quantitative Courses is not unequivocally undesirable, as the institution we study offers these courses for students who do not intend to pursue quantitative majors.

At the same TP – DP margin, we find marginally receiving the Tentative Recommendation to take Pre-Calculus causes more course withdrawals. We also find suggestive evidence these marginal students also perform worse in Calculus. Although this recommendation seems to cause students to substitute Non-Major Quantitative Courses with Calculus, the marginal students do not seem to be benefiting.

However, if we assume that students at the TP – DP margin have equivalent intentions and potential toward quantitative majors (as the logic of the RD design assumes), the fact that there is a statistically and substantially significant estimate at the TP and DP margin suggests that an unintended side effect of the recommendation policy is that the placement recommendations are not matched optimally to the students' quantitative preparation. Either too many DP students are pursuing Non-Major Quantitative Courses, or too few TP students are, or perhaps even both co-occur. Unfortunately, the RD analysis here is agnostic about the way in which the recommendation policy should be adjusted. Such a decision must instead be derived from the educational goals of the institution.

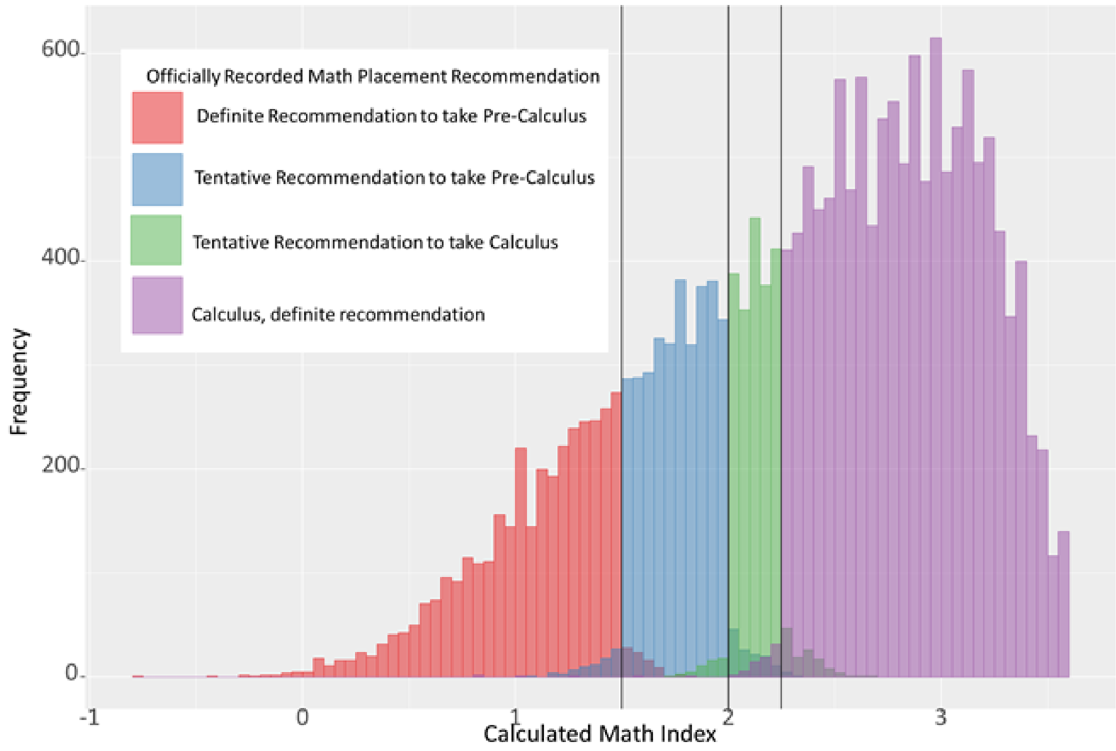
We find statistically and substantially significant effects on cumulative attempted credits in Math, Statistics, and Economics in the second semester. Students who are marginally given the lowest recommendation, a Definite Recommendation to take Pre-Calculus, take half as many Economics credits by their second semester. We find limited evidence this difference persists.

Additional findings of limited evidence on Biology and Chemistry course-taking are puzzling juxtaposed with statistically and substantially significant results on ever taking Calculus. Are students induced to take Calculus not using the Calculus experience for future coursework? It could be that induced Calculus course-taking is associated with detailed course-taking patterns we cannot measure by grouping courses by subject. Alternatively, it is possible that the experience of taking Calculus for those students affected by the recommendation does not encourage–or even discourages–further coursework dependent on Calculus. If policy makers want students to go beyond taking Calculus, evidence suggests that it is insufficient to place students at FAC into Calculus courses, at least those students at the level of preparation at the recommendation margin. Future research on students' motivations to take Calculus or on their actual experience taking Calculus could reveal more potent policy levers.

Table 3.1: Sample Selection

	Female	White	Black	Asian	Hisp	ACT Math	ACT Reading	SAT Math	SAT Verbal	HS GPA
		Proportion of				ACT Scores		SAT Scores		HS
	Female	White	Black	Asian	Hisp	Math	Reading	Math	Verbal	GPA
Starting Sample										
mean	0.509	0.658	0.057	0.136	0.051	28.599	28.732	671.408	630.08	3.742
sd	(0.5)	(0.474)	(0.232)	(0.343)	(0.221)	(4.038)	(4.443)	(77.16)	(78.501)	(0.281)
N	35250	35250	35250	35250	35250	25342	25350	19301	19301	34112
1: Not Incoming to College of Arts and Sciences										
mean	0.371	0.662	0.049	0.142	0.043	29.491	28.363	685.338	622.63	3.707
sd	(0.483)	(0.473)	(0.216)	(0.349)	(0.203)	(4.165)	(4.585)	(78.879)	(83.081)	(0.325)
N	10286	10286	10286	10286	10286	7342	7351	5843	5843	9837
2: Not Incoming Freshmen										
mean	0.523	0.357	0.298	0.115	0.1	23.476	24.443	617.753	577.008	3.315
sd	(0.5)	(0.479)	(0.458)	(0.319)	(0.301)	(4.742)	(5.38)	(104.571)	(86.807)	(0.379)
N	1881	1881	1881	1881	1881	1202	1202	899	899	1601
3: Have Advanced Placement Calculus Credit										
mean	0.517	0.69	0.01	0.197	0.018	31.39	30.408	707.575	659.804	3.865
sd	(0.5)	(0.463)	(0.099)	(0.398)	(0.134)	(2.79)	(3.617)	(57.038)	(67.949)	(0.16)
N	1308	1308	1308	1308	1308	1109	1109	664	664	1307
4: Have Other Calculus Credit										
mean	0.625	0.658	0.05	0.132	0.056	28.155	27.905	656.8	625.6	3.834
sd	(0.485)	(0.475)	(0.219)	(0.339)	(0.23)	(3.243)	(4.375)	(68.877)	(64.209)	(0.194)
N	357	357	357	357	357	296	296	150	150	351
5: Final Sample										
mean	0.571	0.68	0.043	0.132	0.053	28.38	29.139	666.727	636.226	3.781
sd	(0.495)	(0.466)	(0.202)	(0.339)	(0.224)	(3.608)	(4.125)	(72.066)	(73.981)	(0.22)
N	21418	21418	21418	21418	21418	15393	15392	11745	11745	21016

Figure 3.1: Comparing Calculated Math Index with Official Recommendations



NOTES –

Vertical lines represent the cutoff values for receiving different recommendations based on calculated Math Index values. If we perfectly predicted students' recommendations with our calculated Math Index, then there would be no overlap at these vertical lines.

Table 3.2: McCrary Density Test - Differences in the Distribution of Calculated Math Index at Cutoffs for each Incoming Cohort

	2002	2003	2004	2005	2006	2007	2008	Overall
Tentative Recommendation to take Pre-Calculus - Definite Recommendation to take Pre-Calculus (TC – DC)								
Bandwidth	0.251	0.247	0.219	0.187	0.166	0.178	0.167	0.291
Theta	0.04	0.047	0.256	0.37	–0.368	0.258	–0.383	0.035
SE	(0.205)	(0.195)	(0.203)	(0.268)	(0.29)	(0.255)	(0.259)	(0.074)
T-Test	0.197	0.244	1.261	1.382	–1.267	1.013	–1.479	0.477
N	1060	1170	1171	1065	769	882	933	21418
Tentative Recommendation to take Calculus - Tentative Recommendation to take Pre-Calculus (TC – TP)								
Bandwidth	0.071	0.081	0.063	0.097	0.104	0.08	0.077	0.29
Theta	0.556	0.152	0.688	0.222	0.461	0.125	0.748	0.124
SE	(0.415)	(0.386)	(0.408)	(0.255)	(0.308)	(0.317)	(0.365)	(0.065)
T-Test	1.341	0.394	1.685	0.87	1.495	0.394	2.048	1.904
N	754	859	910	865	680	770	789	21418
Definite Recommendation to take Calculus - Tentative Recommendation to take Calculus (DC – TC)								
Bandwidth	0.127	0.183	0.163	0.134	0.149	0.155	0.189	0.323
Theta	–0.243	–0.205	0.329	–0.234	0.217	–0.11	–0.175	–0.054
SE	(0.247)	(0.205)	(0.211)	(0.241)	(0.229)	(0.204)	(0.216)	(0.057)
T-Test	–0.983	–1.003	1.556	–0.974	0.95	–0.542	–0.813	–0.934
N	1869	2100	2160	2425	2165	2321	1328	21418

NOTES – Shown are the bandwidths difference, and standard error of the differences from the McCrary density test. The unrounded calculated math index is used, where the calculated math index used for recommendation is rounded to the nearest hundredths place.

Table 3.3: Proportion of Students who Take Quantitative Courses over Recommendations

	Pre-Calculus By the ··· Semester			Calculus By the ··· Semester			Non-Major Quantitative By the ··· Semester			Sample Size
	2nd	8th	12th	2nd	8th	12th	2nd	8th	12th	
Definitely Take Pre-Calculus (DP)	0.404	0.437	0.44	0.205	0.276	0.279	0.019	0.208	0.221	3623
Tentatively Take Pre-Calculus (TP)	0.358	0.378	0.38	0.366	0.444	0.446	0.017	0.174	0.182	3514
Tentatively Take Calculus (TC)	0.137	0.145	0.146	0.478	0.539	0.542	0.02	0.166	0.177	2136
Definitely Take Calculus (DC)	0.013	0.017	0.017	0.388	0.429	0.43	0.025	0.122	0.129	12145

Notes – This shows the proportion of students who receive different recommendations that take Pre-Calculus, Calculus, or a Non-Major Quantitative Course that counts towards FAC's quantitative course requirement. These are not mutually exclusive groups, and are not all exhaustive possibilities. Students may take quantitative courses that satisfy at least one major's course requirements to satisfy the distributional quantitative course requirement. Students who drop out of FAC before the 2nd, 8th, or 12th semesters have their data carried forward and it is possible students leave never taking any of these courses.

Figure 3.2: Proportion of White Students Over the Math Index

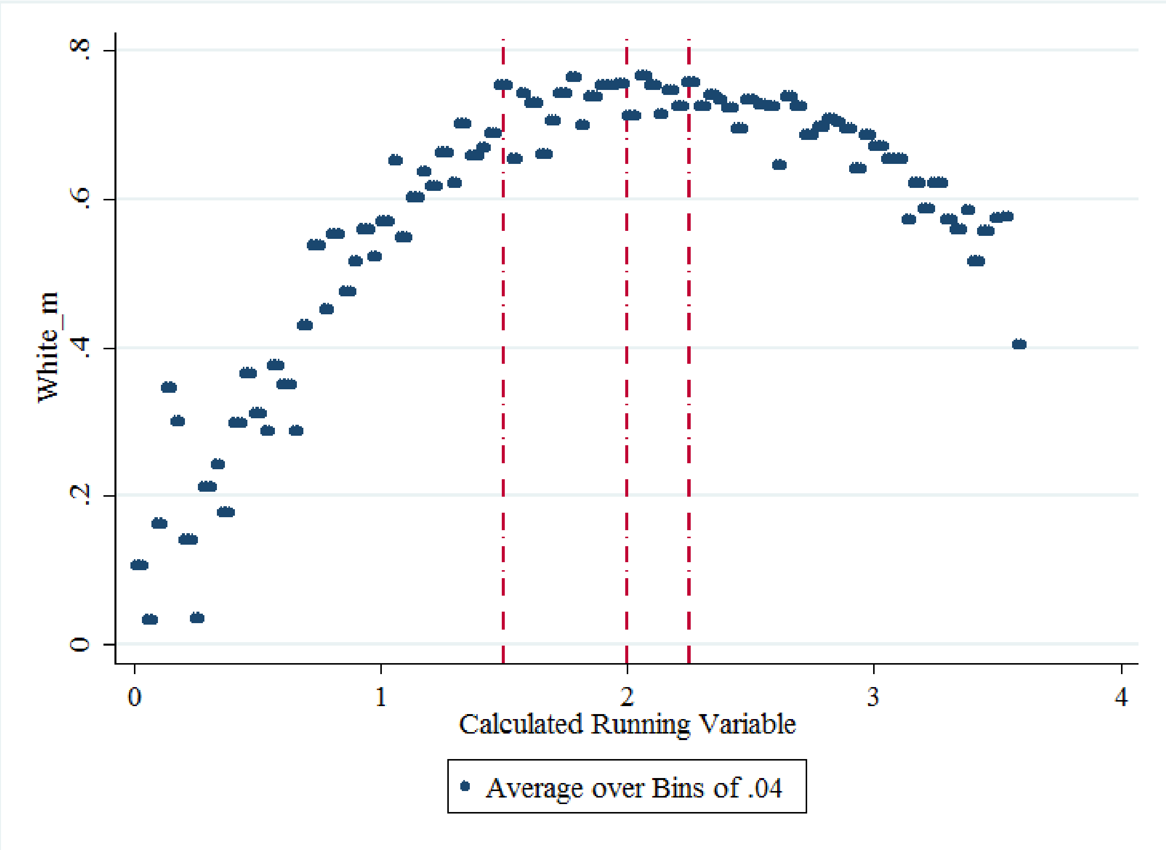


Figure 3.3: Proportion of Black Students Over the Math Index

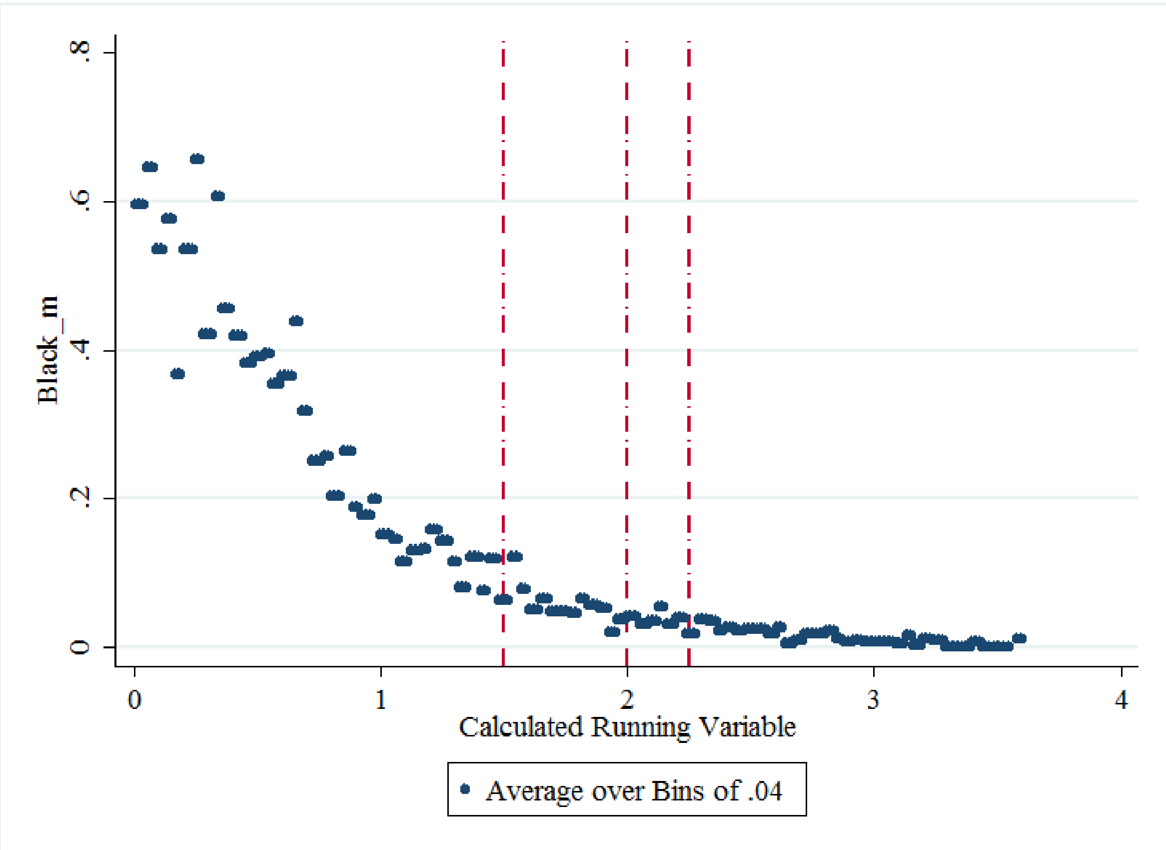


Figure 3.4: ACT Math Score Over the Math Index

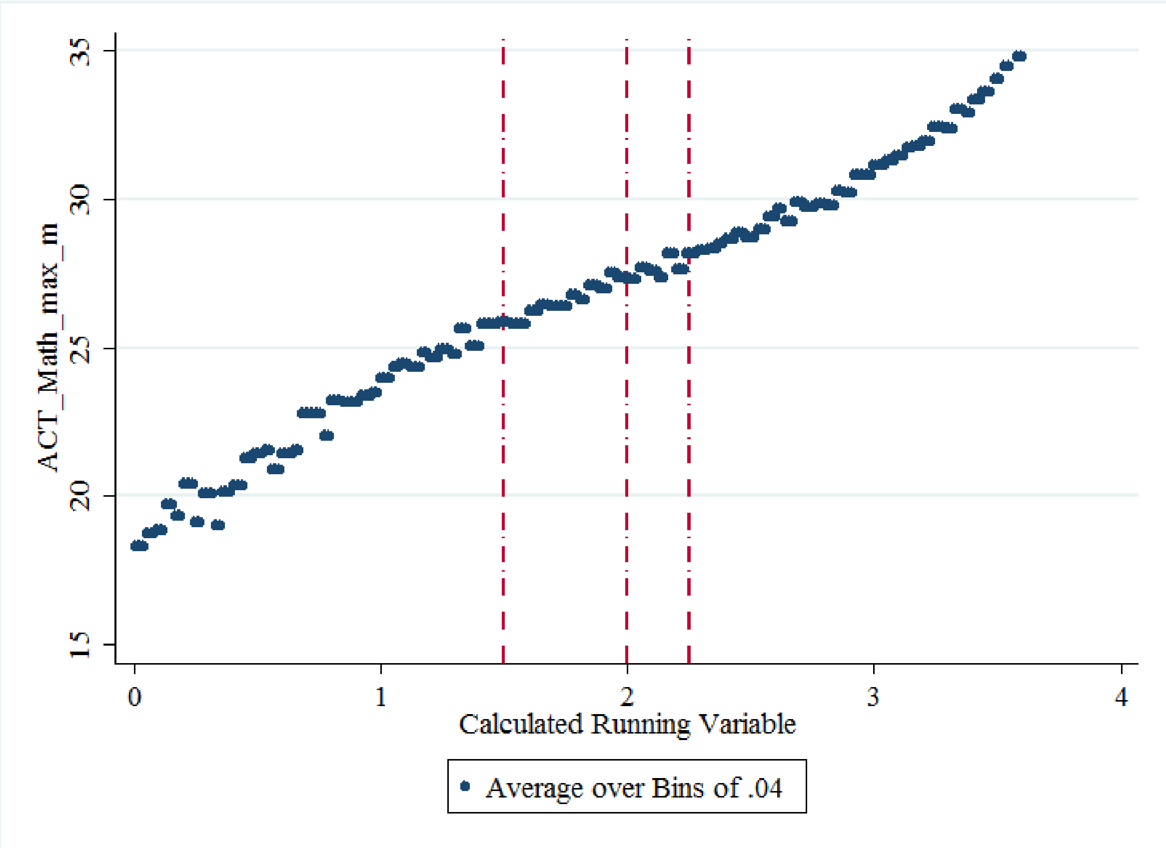


Figure 3.5: SAT Math Score Over the Math Index

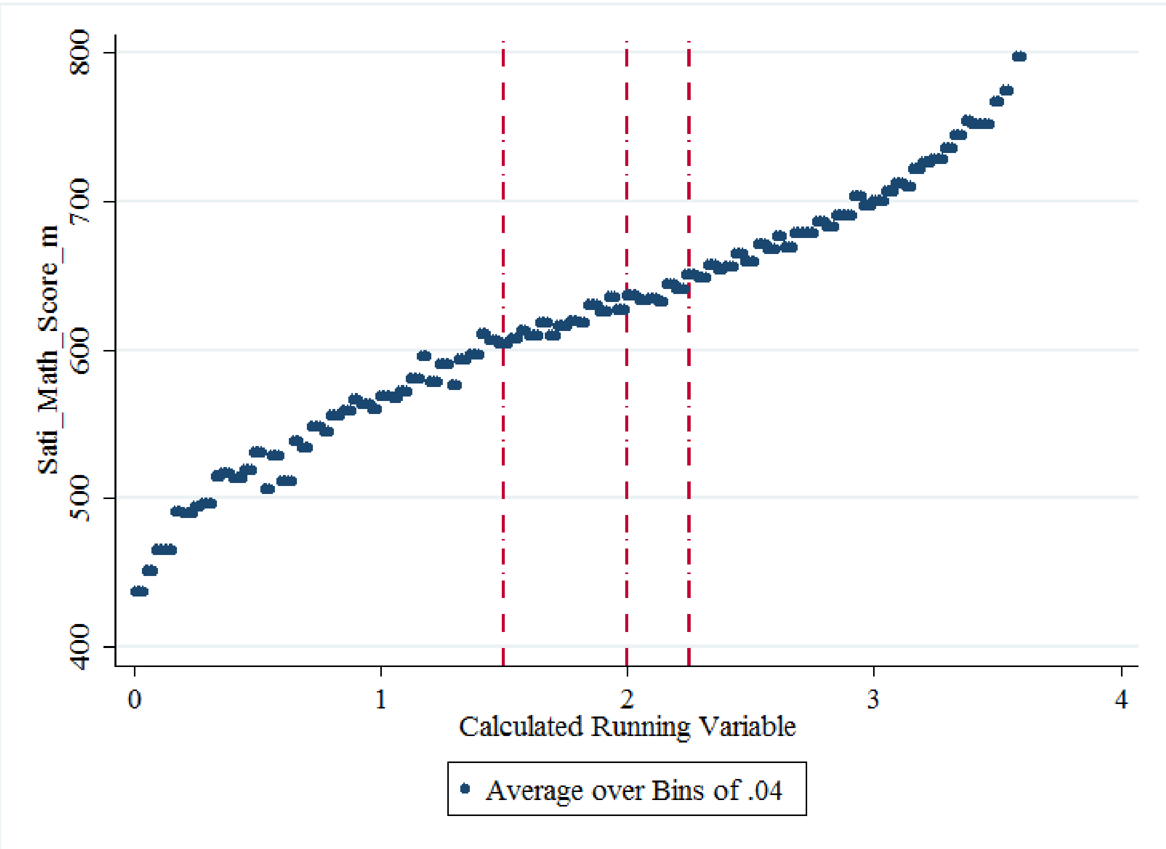


Table 3.4: Balance Test: Non-Parametric Estimates in Student Characteristics

	Reported Race							Enrollment	ACT				SAT		High School
Cutoff @ 2.25	Asian	Black	Hisp	White	Female	Multiple	None	Year	English	Math	Reading	Science	Math	Verbal	GPA
DP – TP															
Estimates	-0.003	-0.052	0.018	-0.013	0.002	-0.002	0.004	-0.346	0.331	-0.056	0.416	0.432	-6.812	-2.188	0.018
P-Values	0.859	0.103	0.430	0.778	0.970	0.932	0.806	0.085	0.455	0.837	0.396	0.302	0.395	0.813	0.526
Bandwidth	0.217	0.118	0.205	0.164	0.217	0.222	0.231	0.185	0.180	0.145	0.186	0.142	0.177	0.232	0.177
Control Mean	0.036	0.096	0.053	0.753	0.608	0.047	0.031	2005.100	27.443	26.005	27.891	25.741	610.196	616.636	3.698
Interval Size	3031	1706	2937	2377	3031	3200	3380	2651	2035	1644	2034	1549	1104	1469	2457
TP – TC															
Cutoff @ 2	Asian	Black	Hisp	White	Female	Multiple	None	Year	English	Math	Reading	Science	Math	Verbal	GPA
Estimates	0.082	0.003	-0.023	-0.084	-0.035	0.001	0.038	-0.129	-0.400	0.001	0.034	0.122	15.333	-8.992	-0.051
P-Values	0.032	0.868	0.454	0.111	0.589	0.946	0.092	0.622	0.530	0.998	0.963	0.800	0.099	0.481	0.125
Bandwidth	0.062	0.114	0.110	0.094	0.079	0.118	0.094	0.085	0.070	0.076	0.070	0.081	0.088	0.094	0.085
Control Mean	0.022	0.037	0.096	0.776	0.660	0.036	0.016	2005.328	28.806	27.495	28.848	26.476	625.003	623.638	3.794
Interval Size	1120	2156	1988	1788	1372	2156	1788	1575	829	1099	829	1190	742	853	1559
TC – DC															
Cutoff @ 2.25	Asian	Black	Hisp	White	Female	Multiple	None	Year	English	Math	Reading	Science	Math	Verbal	GPA
Estimates	0.041	-0.009	0.011	-0.140	-0.018	0.042	0.051	-0.049	0.769	0.848	0.696	1.243	6.661	2.244	-0.107
P-Values	0.199	0.568	0.682	0.012	0.762	0.070	0.089	0.833	0.112	0.014	0.246	0.002	0.468	0.854	0.002
Bandwidth	0.092	0.106	0.090	0.071	0.093	0.061	0.072	0.094	0.083	0.066	0.078	0.085	0.080	0.079	0.052
Control Mean	0.056	0.021	0.049	0.837	0.588	0.000	0.030	2005.005	28.238	27.800	28.804	26.291	645.979	632.992	3.836
Interval Size	1996	2206	1996	1527	1996	1292	1527	1996	1340	1028	1142	1340	867	755	1061

NOTES – : Estimated differences between covariates use the estimator from Calonico et al. (2014), with robust P-Values and bandwidth to account for large bandwidth selection in standard non-parametric RD models. Reported robust p-values, bandwidth, and control mean are reported using the Stata command rdrobust, the accompanying command with the aforementioned work.

Table 3.5: Non-Parametric RD Estimates on Cumulative Credits and Graduation

Cumulative Credits									
	TP – DP			TC – TP			DC – TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	0.231	-1.159	-1.775	0.124	3.840	1.912	0.442	0.471	0.242
P-Value	0.678	0.747	0.619	0.899	0.427	0.656	0.28	0.855	0.921
Bandwidth	0.594	0.456	0.453	0.193	0.263	0.299	0.635	0.695	0.754
Control Mean	27.726	95.806	68.419	28.692	93.364	65.949	28.398	94.459	65.965
Interval Size	3444	2648	2648	1556	2131	2303	5673	6319	6815
Cumulative Course Withdrawals									
	TP – DP			TC – TP			DC – TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	-0.082	0.568	0.594	-0.033	-0.169	-0.009	-0.136	0.182	0.211
P-Value	0.430	0.046	0.012	0.771	0.439	0.965	0.206	0.585	0.438
Bandwidth	0.458	0.461	0.499	0.256	0.331	0.267	0.214	0.201	0.220
Control Mean	0.330	0.853	0.547	0.205	0.952	0.673	0.303	0.839	0.610
Interval Size	2648	2799	2871	1954	2586	2131	1951	1951	1951

Graduated by the 8th Semester			
	TP – DP	TC – TP	DC – TC
	8th	8th	8th
Estimate	-0.044	0.024	-0.006
P-Value	0.074	0.519	0.819
Bandwidth	0.613	0.281	0.364
Control Mean	0.965	0.927	0.933
Interval Size	3568	2303	3402

NOTES – : ITT estimates compare the marginal student between different recommendations, abbreviated as the Definite Recommendation to take Pre-Calculus (DP), Tentative Recommendation to take Pre-Calculus (TP), Tentative Recommendation to take Calculus (TC) and Definite Recommendation to take Calculus (DC). RD estimates use the estimator from Calonico et al. (2014), with robust P-Values and bandwidth to account for large bandwidth selection in standard non-parametric RD models.
For all following tables, we show p-values, bandwidths, control means, and Interval Sizes that are robust to the inherent large bandwidth selection problem. These statistics are reported using the Stata command rdrobust, the accompanying command with the aforementioned work.
“ 8th - 2nd” is where the outcome is the difference between 8th and 2nd semester outcomes.

Table 3.6: Non-Parametric RD Estimates on Ever Taking Pre-Calculus, Calculus, and a "Non-Major Quantitative Course"

Ever Take Pre–Calculus

	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	-0.006	-0.02	-0.022	-0.078	-0.079	-0.004	0.016	0.028	0.001
P-Value	0.91	0.698	0.394	0.267	0.282	0.772	0.698	0.513	0.951
Bandwidth	0.566	0.656	0.388	0.237	0.228	0.242	0.227	0.228	0.362
Control Mean	0.448	0.477	0.044	0.306	0.312	0.008	0.073	0.076	0.009
Interval Size	3329	3799	2326	1859	1859	1954	2125	2125	3402

Ever Take Calculus

	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	0.165	0.183	-0.008	0.147	0.076	-0.045	-0.023	0.004	0.023
P-Value	0.011	0.007	0.771	0.03	0.263	0.241	0.733	0.955	0.357
Bandwidth	0.369	0.377	0.556	0.297	0.302	0.225	0.251	0.26	0.315
Control Mean	0.226	0.273	0.073	0.383	0.473	0.073	0.598	0.63	0.035
Interval Size	2183	2183	3214	2303	2461	1859	2284	2420	2787

Ever Take Non-Major Quantitative Course

	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	-0.023	-0.129	-0.098	0.009	-0.019	-0.026	0.007	-0.023	-0.025
P-Value	0.036	0.003	0.013	0.655	0.731	0.675	0.666	0.624	0.605
Bandwidth	0.563	0.588	0.709	0.223	0.271	0.231	0.243	0.273	0.242
Control Mean	0.022	0.264	0.237	0.015	0.181	0.169	0.012	0.168	0.154
Interval Size	3329	3444	4154	1859	2131	1859	2284	2420	2284

Table 3.7: Non-Parametric RD Estimates on First Earned Pre-Calculus and Calculus Grades

First Earned Pre-Calculus Grade			
	TP – DP	TC – TP	DC – TC
Estimates	0.080	0.077	0.020
P-Values	0.313	0.162	0.599
Bandwidth	0.235	0.111	0.071
Control Mean	3.344	3.733	3.942
Interval Size	3359	2153	1526
First Earned Calculus Grade			
	TP – DP	TC – TP	DC – TC
Estimates	-0.372	-0.199	-0.046
P-Values	0.007	0.076	0.715
Bandwidth	0.107	0.109	0.074
Control Mean	3.500	3.309	3.031
Interval Size	1527	1965	1514

Table 3.8: Non-Parametric RD Estimates on Taking a Course that Requires Calculus and Cumulative Chemistry and Biology Credits

Ever Take Course that Requires Calculus

	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	0.008	0.044	0.027	0.002	0.075	0.076	-0.008	0.041	0.036
P-Value	0.641	0.386	0.565	0.963	0.258	0.162	0.857	0.511	0.463
Bandwidth	0.491	0.358	0.367	0.311	0.221	0.221	0.247	0.244	0.274
Control Mean	0.017	0.13	0.118	0.073	0.147	0.078	0.114	0.288	0.176
Interval Size	2871	2030	2183	2461	1859	1859	2284	2284	2420

Cumulative Chemistry Credits

	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	-0.204	0.035	0.378	0.143	0.377	0.324	-0.177	-1.036	-0.806
P-Value	0.428	0.965	0.53	0.674	0.709	0.667	0.696	0.372	0.345
Bandwidth	0.667	0.475	0.447	0.328	0.266	0.265	0.208	0.209	0.221
Control Mean	1.446	3.346	1.854	1.641	4.322	2.638	1.930	5.440	3.437
Interval Size	3971	2799	2648	2586	2131	2131	1951	1951	2125

Cumulative Biology Credits

	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	0.254	0.564	0.299	-0.18	0.518	0.76	-0.123	-0.798	-0.766
P-Value	0.235	0.439	0.647	0.481	0.577	0.351	0.689	0.395	0.381
Bandwidth	0.581	0.499	0.474	0.311	0.254	0.255	0.258	0.267	0.239
Control Mean	0.823	3.568	2.766	1.046	3.97	2.874	1.209	5.181	4.116
Interval Size	3444	2871	2799	2461	1954	1954	2284	2420	2125

Table 3.9: Non-Parametric RD Estimates on Cumulative Credits in Math, Statistics, and Economics

Cumulative Math Credits	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	0.718	0.236	-0.274	-0.172	-0.635	-0.555	-0.005	0.146	0.149
P-Value	0.083	0.625	0.266	0.665	0.32	0.301	0.987	0.788	0.674
Bandwidth	0.423	0.5	0.594	0.314	0.321	0.257	0.276	0.31	0.365
Control Mean	2.49	3.428	0.874	3.094	4.301	1.239	2.917	3.787	0.935
Interval Size	2503	2871	3444	2461	2586	1954	2420	2787	3402
Cumulative Statistics Credits									
	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	0.279	0.383	0.204	-0.029	0.089	0.013	0.497	0.005	-0.252
P-Value	0.123	0.257	0.469	0.907	0.828	0.967	0.053	0.989	0.519
Bandwidth	0.456	0.444	0.526	0.239	0.251	0.333	0.204	0.269	0.199
Control Mean	0.434	2.127	1.663	0.742	2.534	1.848	0.674	2.869	2.064
Interval Size	2648	2648	3136	1859	1954	2586	1951	2420	1766
Cumulative Economics Credits									
	TP – DP			TC – TP			DC - TC		
	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd	2nd	8th	8th - 2nd
Estimate	0.703	1.152	0.361	-0.133	0.664	1.269	0.212	0.179	0.017
P-Value	0.014	0.353	0.726	0.695	0.511	0.178	0.455	0.874	0.988
Bandwidth	0.323	0.346	0.371	0.233	0.27	0.238	0.242	0.308	0.27
Control Mean	0.629	3.037	2.400	1.289	3.677	1.978	1.126	5.142	4.176
Interval Size	1951	2030	2183	1859	2131	1859	2284	2787	2420

Table 3.10: Non-Parametric RD Estimates on Graduating in Selected Majors by the Eighth Semesters

Graduate in Biology			
	TP – DP 8th	TC – TP 8th	DC - TC 8th
Estimate	-0.019	0.001	-0.044
P-Value	0.531	0.986	0.369
Bandwidth	0.445	0.336	0.2
Control Mean	0.069	0.077	0.116
Interval Size	2648	2586	1766
Graduate in Economics			
	TP – DP 8th	TC – TP 8th	DC - TC 8th
Estimate	0.054	-0.005	0.003
P-Value	0.131	0.872	0.925
Bandwidth	0.348	0.306	0.314
Control Mean	0.034	0.065	0.092
Interval Size	2030	2461	2787
Graduate in Physics, Math, or Statistics			
	TP – DP 8th	TC – TP 8th	DC - TC 8th
Estimate	0.000	-0.006	0.005
P-Value	0.998	0.709	0.723
Bandwidth	0.706	0.225	0.346
Control Mean	0.008	0.014	0.015
Interval Size	4154	1859	3184

Bibliography

Daniel Ackerberg. Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. *International Economic Review*, 44:1007–1040, 2003.

Victor Aguirregabira and Pedro Mira. Dynamic Discrete Choice Structural Models: A Survey. *Journal of Econometrics*, 156:38–67, 2010.

Ingvild Alms, Alexander W. Cappelen, Kjell G. Salvanes, Erik . Srensen, and Bertil Tungodden. What Explains the Gender Gap in College Track Dropout? Experimental and Administrative Evidence. *American Economic Review*, 106(5):296–302, 2016.

Joseph Altonji and Charles Pierret. Employer Learning and Statistical Discrimination. *The Quarterly Journal of Economics*, 116, 2001.

Joseph Altonji, Peter Arcidiacono, and Aranud Maurel. The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. In Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, editors, *Handbook of Education*, chapter 7, pages 305–392. Elsevier North-Holland, 2015.

Joseph G Altonji. The Demand for and Return to Education When Education Outcomes are Uncertain. *Journal of Labor Economics*, 11(1):48–83, 1993.

Joseph G. Altonji, Erica Blom, and Costas Meghir. Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics*, 4:185 – 223, 2012.

Rodney Andrews, Jing Li, and Michael Lovenheim. Heterogeneous Paths Through College: Detailed Patterns and Relationships with Graduation and Earnings. Working Paper 19935, National Bureau of Economic Research, February 2014.

Massimo Anelli and Giovanni Peri. Gender of Siblings and Choice of College Major. *CESifo Economic Studies*, 61(1):53–71, 2015.

Massimo Anelli and Giovanni Peri. The Effects of High School Peers' Gender on College Major, College Performance and Income. *CESifo Working Paper No. 6014*, 2016.

Peter Arcidiacono. Ability Sorting and the Returns to College Major. *Journal of Econometrics*, 121:343–375, 2004.

Peter Arcidiacono and Paul Ellickson. Practical Methods for the Estimation of Dynamic Choice Models. *Annual Review of Economics*, 3:363–394, 2011.

Peter Arcidiacono and Michael Lovenheim. Affirmative Action and the Quality-Fit Tradeoff. *NBER Working Paper No. 20962*, 2015.

Peter Arcidiacono, Esteban Aucejo, and Ken Spenner. What Happens After Enrollment? An Analysis of the Time Path of Racial Differences in GPA and Major Choice. *IZA Journal of Labor Economics*, 1, 2012a.

Peter Arcidiacono, V Joseph Hotz, and Songman Kang. Modeling College Major Choices using Elicited Measures of Expecations and Counterfactuals. *Journal of Econometrics*, 166: 3–16, 2012b.

Peter Arcidiacono, V. Joseph Hotz, Arnaud Maurel, and Teresa Romano. Recovering Ex Ante Returns and Preferences for Occupations using Subjective Expectations Data. *NBER Working Paper Series No. 20626*, 2014.

Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College Attrition and the Dynamics of Information Revelation. *Journal of Political Economy*, 2016.

Christopher Avery, Oded Gurantz, Michael Hurwitz, and Jonathan Smith. Shifting College Majors in Response to Advanced Placement Exam Scores. *NBER Working Paper Series No. 22841*, 2016.

Rachel Baker, Eric Bettinger, Brian Jacob, and Ioana Marinescu. The Effect of Labor Market Information on Community College Students Major Choice. *NBER Working Paper Series No. 23333*, 2017.

Talia Bar, Vrinda Kadiyali, and Asaf Zussman. Grade Information and Grade Inflation: The Cornell Experiment. *Journal of Economic Perspectives*, 23:93–108, 2009.

Alan Barreca, Jason Lindo, and Glen Waddell. Heaping-Induced Bias in Regression Discontinuity Designs. *Economic Inquiry*, 54(1):268–293, 2016.

Magali Beffy, Denis Fougere, and Arnaud Miguel. Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? *The Review Of Economics And Statistics*, 94(1):334–347, 2012.

Eric Bettinger. American Universities in a Global Market - To Be or Not To Be: Major Choices in Budding Scientists. In Charles T Clotfelter, editor, *American Universities in a Global Market*, pages 69–98. University of Chicago Press, 2010.

Eric Bettinger and Bridget Terry Long. Remediation at the Community College: Student Participation and Outcomes. *New Directions for Community College*, 129:17–26, 2005.

Eric P. Bettinger and Bridget Terry Long. Addressing the Needs of Underprepared Students in Higher Education. *Journal of Human Resources*, 44(3):736–771, 2009.

Dan Black, Seth Sanders, and Lowell Taylor. The Economic Reward for Studying Economics. *Economic Inquiry*, 41:365–377, 2003.

Angela Boatman and Bridget Terry Long. Does Remediation Work for All Students? How the Effects of Postsecondary Remedial and Developmental Courses Vary by Level of Academic Preparation. An NCPR Working Paper. *National Center for Postsecondary Research*, 2010.

Paolo Bordon and Chao Fu. College-Major Choice to College-Then-Major Choice. *The Review of Economic Studies*, 82(4):1247–1288, 2015.

Charles Brown and Mary Corcoran. High School Courses, College Majors and Sex Differences in Wages. *Journal of Labor Economics*, 15:431–465, 1997.

Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326, 2014.

Emily Forrest Cataldi, Caitlin Green, Robin Henke, Terry Lew, Jennie Woo, Bryan Shepherd, Peter, and Siegel. 200809 Baccalaureate and Beyond Longitudinal Study (B&B:08-09): First Look. Technical report, U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2011. Retrieved 7-July-2016 from http://nces.ed.gov/pubsearch .

Xianglei Chen and Thomas Weko. Students Who Study Science, Technology, Engineering, and Mathematics (STEM) in Postsecondary Education. Technical report, National Center for Education Statistics, 2009.

Thomas Cook. Waiting for Life to Arrive: A History of the Regression-Discontinuity Design in Psychology, Statistics and Economics. *Journal of Econometrics*, 142(2):636–654, 2008.

Gregory Crawford and Matthew Shum. Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, 73:1137–73, 2005.

Lisandro Dalcin, Rodrigo Paz, and Mario Storti. MPI for Python, Journal of Parallel and Distributed Computing. *Journal of Parallel and Distributed Computing*, 65:1108–1111, 2005.

Rajeev Darolia and Cory Koedel. How High Schools Explain Students Initial Colleges and Majors. *Working Paper - WP16-09*, 2016.

Morris H DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1979.

Eleanor Wiske Dillon and Jeffrey A. Smith. The Consequences of Academic Match between Students and Colleges. *CESifo Working Paper Series No. 6344*, 2017.

John C Gittins. Bandit Processes and Dynamic Allocation Indices (with Discussion). *Journal of the Royal Statistics Society: Series B*, 41(2), 1979.

Dan Goldhaber, James Cowan, Mark C. Long, and Nick Huntington-Klein. College Curricular Dispersion: More Well Rounded or Less Well Trained? *National Center for Analysis of Longitudinal Data in Education Research Working Paper 130*, 2015.

Michael Gottfried and Robert Bozick. Supporting the STEM Pipeline: Linking Applied STEM Course-Taking in High School to Declaring a STEM Major in College. *Education Finance and Policy*, 11(2):177–202, 2016.

Amanda Griffith. Persistence of Women and Minorities in STEM Field Majors: Is It the School that Matters? *Economics of Education Review*, 29:911–922, 2010.

Amanda Griffith. Faculty Gender in the College Classroom: Does it Matter for Achievement and Major Choice? *Southern Economic Journal*, 81(1):211–231, 2014.

Justine Hastings, Chrisopher Nielsen, and Seth Zimmerman. Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile. *NBER Working Paper Series No. 19241*, 2013.

James Heckman and Burton Singer. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, 52:271–320, 1984.

Steven Hemelt. The College Double Major and Subsequent Earnings. *Education Economics*, 18:167–189, 2010.

Lutz Hendricks and Oksana Leukina. How Risky is College Investment? *CESifo Working Paper*, 2015.

Mark Hoekstra. The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach. *The Review of Economics and Statistics*, 91(4):717–724, 2009.

Catherine Horn, Zoë McCoy, Lea Campbell, and Cheryl Brock. Remedial Testing and Placement in Community Colleges. *Community College Journal of Research and Practice*, 33(6): 510–526, 2009.

Julian Hsu. Do Grades Matter? Evidence from College Transcripts. *Working Paper*, 2017.

Mark Israel. Services as Experience Goods: An Empirical Examination of Consumer Learning in Automobile Insurance. *American Economic Review*, 95:1444–1463, 2005.

David Jaegar and Marianne Page. Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education. *The Review of Economics and Statistics*, 78(4):733–740, 1996.

Jonathan James. Ability Matching and Occupational Choice. *Cleveland Federal Reserve Working Paper*, 2011.

Eric Jones, Travis Oliphant, Pearu Peterson, and et al. SciPy: Open source scientific tools for Python, 2001. URL http://www.scipy.org/. [Online; accessed 2016-04-26].

Thomas J. Kane and Cecilia Elena Rouse. Labor-Market Returns to Two-and Four-Year College. *The American Economic Review*, 1995.

Michael Keane and Kenneth Wolpin. Career Decisions of Young Men. *Journal of Political Economy*, 105:473–522, 1997.

Josh Kinsler and Ronni Pavan. The Specificity of Human Capital: Evidence from College Major Choice. *Journal of Labor Economics*, 33:29–61, 2015.

Lars Kirkebøen, Edwin Leuven, and Magne Mogstad. Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics*, 131:1057–1111, 2016.

David Lee. The Electoral Advantage to Incumbency and Voters' Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Elections to the U.S. House. *NBER Working Paper Series 8441*, 2001.

David Lee and Thomas Lemieux. Regression Discontinuity Design in Economics. *Journal of Economic Literature*, 48:281–355, 2009.

Mark C Long, Dan Goldhaber, and Nick Huntington-Klein. Do Students' College Major Choices Respond to Changes in Wages? *Center For Education Data & Research Working Paper 2014-6*, 2014.

Joyce Main and Ben Ost. The Impact of Letter Grades on Student Course Selection and Major Choice: Evidence from a Regression-Discontinuity Design. *The Journal of Economic Education*, 45:1–10, 2014.

Ofer Malamud. Breadth vs. Depth: The Timing Of Specialization In Higher Education. *Labour*, 24(4):359–390, 2010.

Ofer Malamud. Discovering One's Talent: Learning from Academic Specialization. *Industrial & Labor Relations Review*, 64(375), 2011.

Charles A Malgwi, Martha A Howe, and Priscilla A Burnaby. Influences on Students' Choice of College Major. *Journal of Education For Business*, 80(5):275–282, 2005.

Paco Martorell and Isaac McFarlin. Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes. *The Review of Economics and Statistics*, 93(2): 436–454, 2011.

Justin McCrary. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714, 2008.

Daniel McFadden. Modelling the Choice of Residential Location. *Spatial Interaction Theory And Planning Models*, pages 75–96, 1978.

Tatiana Melguizo, Johannes M. Bos, Federick Ngo, Nicholas Mills, and George Prather. Using a Regression Discontinuity Design to Estimate the Impact of Placement Decisions in Developmental Math. *Research In Higher Education*, 57(2):123–151, 2015.

Claude Montmarquette, Kathy Cannings, and Sophie Mahseredjian. How Do Young People Choose Majors? *Economics of Education Review*, 21:543–556, 2002.

Ben Ost. The Role of Peers and Grades in Determining Major Persistence in the Sciences. *Economics of Education Review*, 29:923–934, 2010.

Ann L. Owen. Grades, Gender, and Encouragement: A Regression Discontinuity Analysis. *Journal of Economic Education*, 41:217–234, 2008.

John P Papay, Richard J Murnane, and John B Willett. How Performance Information Affects Human-Capital Investment Decisions: The Impact of Test-Score Labels on Educational Outcomes. *NBER Working Paper Series No. 17120*, 2011.

Joshua Price. The Effect of Instructor Race and Gender on Student Peristence in STEM Fields. *Education Of Economics Review*, 29:901–910, 2010.

Joseph Robinson. Evaluating Criteria for English Learner Reclassification: A Causal-Effects Approach Using a Binding-Score Regression Discontinuity Design With Instrumental Variables. *Educational Evaluation and Policy Analysis*, 33(3):267–292, 2011.

Alison Rossi and Joni Hersch. The Private and Social Benefits of Double Majors. *Journal of Benefit-Cost Analysis*, 7:292–325, 2016.

Judith Scott-Clayton and Olga Rodriguez. Development, Discouragement, or Diversion? New Evidence on the Effects of College Remediation Policy. *Education Finance and Policy*, 10 (1):4–45, 2015.

Judith Scott-Clayton, Peter M Crosta, and Clive R Belfield. Improving the Targeting of Treatment: Evidence From College Remediation. *Educational Evaluation and Policy Analysis*, 36(3):371–393, 2014.

Jonthan Smith, Michael Hurwitz, and Christopher Avery. Giving College Credit Where it is Due: Advanced Placement Exam Scores and College Outcomes. *Journal of Labor Economics*, 35(1):67–147, 2017.

Jamin Speer. The Gender Gap in College Major: Revisiting the Role of Pre-College Factors. *Labour Economics*, 44:69–88, 2017.

Kevin Stange. An Empirical Investigation of the Option Value of College Enrollment. *American Economic Journal: Applied Economics*, 4(11):49–84, 2012.

Todd Stinebrickner and Ralph Stinebrickner. Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model. *Journal of Labor Economics*, 33:601–644, 2014a.

Todd Stinebrickner and Ralph Stinebrickner. A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *The Review Of Economics Studies*, 81(1):426–472, 2014b.

Sarah H. Turner and William C. Bowen. Choice of Major: The Changing (Unchanging) Gender Gap. *Industrial & Labor Relations Review*, 52:289–313, 1999.

U.S. Department of Education. 2003-04 Beginning Postsecondary Students Longitudinal Study, Second Follow-Up (BPS:04/09). Technical report, National Center for Education Statistics, 2016.

Anthony Varnevale, Ban Cheah, and Jeffrey Strohl. Hard Times: College Majors, Unemployment and Earnings: Not All College Degrees Are Created Equal. Technical report, 2013.

Peter Whittle. Restless Bandits: Activity Allocation in a Changing World. *A Celebration of Applied Probability: Journal of Applied Probability Special Volume*, 25:287–298, 1988.

Matthew Wiswall and Basit Zafar. Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies*, pages 791–824, 2013.

Basit Zafar. Double Majors: One for Me, One for Mom and Dad? *Economic Inquiry*, 50(2): 287–308, 2012.

APPENDIX A

Appendix for Chapter One

A.1 Administrative Data Details

A.1.1 Nature of Data Storage

Administrative data at the University is stored in such a way that information such as ACT and SAT scores is overwritten if students retake them after entering the University. This may be because the student wants to transfer out of the University. I cannot observe these changes, as I can only use a “snapshot” of the data. Very few students transfer out of the University to other institutions.

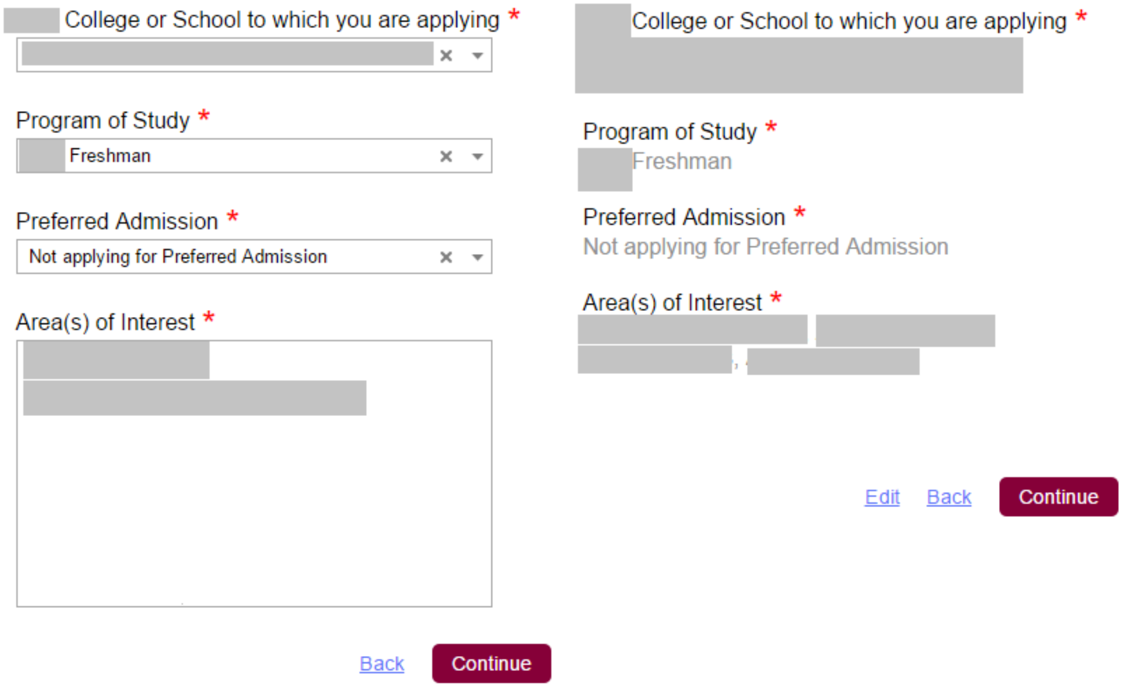
A.1.2 Student’s Pre-College Academic Interests

Student’s pre-college interests are collected by the University from a diverse set of data sources. This includes the questionaire on the SAT, a student profile when registering for the ACT, and the Common Application. The SAT allows students to list up to three majors they are interested in, regardless of whether they are offered at the University. The Common Application allows students to list many more “Areas of Interest,” which are restricted to the majors available at the University, regardless of whether the student is applying to CALS or other colleges at the University. The ACT allows students to list one intended major as well as intended occupation or career.

Below are the prompts from the SAT, ACT, and Common Application students see when listing interests.

> **SAT questionnaire:**
> Items 20. to 22. Choice of Majors – Supply up to three choices of majors to possibly pursue in college.
> **ACT, “Your Plans for the Future”:**
> The college major (program of study) you plan to enter: Your choice of occupation (vocation):

Common Application Area(s) of Interest



A.1.3 Conceptual Issues with Merging on Earnings from External Datasets

One way to collect pecuniary data is to merge on earnings and employment data for observably similar students from highly selective public institutions from nationally representative surveys such as the American Community Survey or NLSY97. However, this encounters two major issues. First, imputation based on observed characteristics leaves little to no individual variation for estimation (consider the case where I use the average earnings conditional on observed characteristics). Second, students who attended college in 2002 to 2011 likely have not yet reached a steady earnings trajectory. If the relevant post-graduation outcome is expected earnings, then it is dangerous to extrapolate from current earnings data. Using historical earnings data does not account for more recent shifts in the economy that not only affect the earnings potentials of different occupations, but the correspondence between occupations and majors.

A.1.4 Coding Specific Major Requirements

The University provides on their website the history of all majors' course requirements. The requirements each student faces depends on their incoming semester. Requirements are tailored for each student, depending on their course history. For example, suppose that either Psychology 101 or Psychology 102 can be taken to satisfy the Psychology major's introductory course requirement. Then I only code the first instance of taking Psychology 101 or Psychology 102 as satisfying the Psychology major's course requirements.

Courses with at least a grade of "C" satisfy majors' course requirements. If a student "fails" on this margin and retakes the course, I only code the first taken course as satisfying the course requirement. Using this measure, I would incorrectly infer a student completes the requirements for a major if she received a grade lower than “C” in the last course she took in that major. Withdrawals are only recorded if the student withdrew at least three weeks into the course. Courses usually schedule the first exam to be around this time. My measurement captures whether or not the student repeats the course because she received below a “C”. If the student retakes the course and passes, likely she will continue taking courses that I measure as making progress.

Some requirements are taking a set of courses instead of individual courses. For example, one requirement in the Physics major is to complete one out of four different sets of courses, where the sets do not overlap. This set is a combination of different lecture and laboratory courses. To code this requirement, I only count the first completed set. If the student does not complete any set, I do not measure any progress in that major. There are similar requirements in other natural science majors, and the Psychology major. My decision implies that I will under measure how students complete majors’ course requirements.

Prerequisite courses are coded as making progress in these majors, and I disregard whether an individual student has satisfied the prerequisite before taking other courses in the major. There is variation in how departments enforce prerequisites, ranging from loosely “advised” prerequisites to strictly preventing students from registering unless they are concurrently taking or previously took the prerequisite. Individual instructors can override these requirements, though these are also subject to individual departmental policies.

At the University, the same course generally cannot be used to satisfy a general distributional requirement and a major course’ requirement. For example, Economics 101 satisfies the general quantitative requirement, but also is required for the Economics major. This means a student who intends to major in Economics must take another course to fulfill the general quantitative requirement. Although several advanced Economics courses also satisfy the general quantitative requirement, the general rule is that a student cannot use one of them to satisfy both Economics and general quantitative requirements.

To avoid potential overlap between majors’ course requirements and general requirements, courses that satisfy distributional requirements span many different majors. For example, courses in Astronomy also satisfy the general quantitative requirement. In certain cases, these rules might place extreme constraints on the students – for example, graduating within a certain time frame. Academic advisors can override these rules, and certain courses can satisfy multiple majors’ requirements if the student is a double major. These decisions are jointly discussed and made with the student.

In this case, the student and academic advisor designate one major to be the “primary major,” and a course that satisfies a general graduation requirement can also satisfy the major requirement in a “secondary major.” Under this multiple major paradigm, the student is able to, on a case-by-case basis, have one course satisfy multiple majors' requirements. There are some un-avoidable overlaps, such as requiring Statistics 101 for most Social Science majors. The student and academic advisor can strategically designate which is the "primary major," which is not observed in the administrative data.

Finally, these requirements are recorded in varying levels of detail. The larger majors such as Psychology, Economics, and Political Science, and scientifically rigorous majors such as Biology, Mathematics, and Computer Science have detailed requirements. They list the specific courses and students can take to satisfy different major requirements. Coding up these majors is relatively straightforward.

However, other smaller majors such as Art History and English have more ambiguous requirements. For example, describing a list of courses to be approved by an advisor, or courses "with an emphasis on Latin American history." To code these requirements would introduce massive measurement error in how progress can be used to infer intention to complete a major. At a minimum, pre-requisite courses are listed in relatively more detail. However, coding only the pre-requisite courses will inflate the measurement error in progress towards completion. This was particularly egregious for certain smaller majors (with less than two hundred graduates over the entire sample period). Aggregating these majors and representing progress in this major group using the maximum progress helps address measurement error in each individual major.

A.1.5 Availability of Pre-Medical and Remedial Courses

The University does not offer a premedical major. The University provides students interested in applying to medical College with the exact courses that satisfy these requirements, and these courses also count towards different major requirements. CALS offers majors in related fields such as Biology and Chemistry. Medical Colleges in the United States require applicants to take courses in Biology, Chemistry, Physics, and Calculus.

The University does not offer any remedial or developmental courses. These are courses designed for under-prepared college students, and cover basic quantitative and reading skills, and do not satisfy any graduation course requirements. The University offers summer programs for these students, and has special versions of Pre-Calculus courses for these students as well. All versions of Pre-Calculus satisfy graduation requirements.

A.1.6 Calendar System at the University – Aggregating to Four Years

The University runs on a semester calendar system, offering standard Fall and Spring semesters with two shortened Summer semesters. Although the transcript data shows students' courses by semester, I use years to simplify how students' choice histories across each major group evolve over time. These decisions also help with the backwards induction calculation to solve the value function – discussed in Subsection A.4.3. This aggregates students' decisions at the start of each semester within a year. Observing that most students leave the University after their fourth years in Table A.1, I model students' decisions over four years, with the fourth year representing all decisions from their fourth year onwards.

Table A.1 shows that very few students are taking Summer courses. Analyzing semester-to-semester creates an issue of how to account for Summer semesters. Summer semesters are not equivalent to Fall and Spring semesters. Summer course-taking is around one-fifth to one-third of Fall semester course-taking. Keeping Summer semesters separate requires accounting for the decision of whether to take Summer courses, while grouping them with the Fall and Spring semester artificially inflates course-taking. I instead look at yearly course-taking. Although students who take Summer semester courses will take more courses per year, I forgo explicitly model whether students decide to take any summer courses instead of other activities such as interning.

I see a large drop of attending students after their fourth year. Assuming that all students who leave before their fifth Fall graduate, this is slightly below the median graduation time of 52 months.[1]

A.1.7 "Non-Requirement Major Courses" that Do Not Satisfy Any Major Groups' Course Requirements

Table A.3 shows a substantial amount of course-taking does not satisfy any major groups' course requirements, around 2% to 6% for the Natural and Social Sciences, and 4% to 10% for Humanities. I find descriptive evidence that taking "Non-Requirement Major Courses" related to different majors is independent of, or at least not a substitute for, making progress across major groups. For this reason, and to reduce the number of state variables, I do not explicitly account for them in the dynamic course-taking model.

By assuming that "Non-Requirement Major Courses" do not affect students' learning about major group match quality, I will over-estimate the effect of information from courses that do satisfy major groups' course requirements. Figure A.2 shows that the distribution of grades between courses that fulfill and do not fulfill major groups' requirements are similar. "Non-Requirement Major Courses" tend to have more grade inflation and bunching at certain grade values. This comes at the cost of tracking students' progress in different major groups over time. To appropriately model taking these "Non-Requirement Major Courses" introduces another level of choice-making, as students decide, for each major group, whether to take courses that satisfies course requirements, and whether to take courses that are related but do not satisfy course requirements. It is likely students are taking these "Non-Requirement Major Courses" for their consumption and learning value.

[1] From the 2007-08 cohort of students starting at public four-year institutions (Cataldi et al., 2011).

Table A.2 finds limited evidence that whether a student takes any "Non-Requirement Major Course" is related to whether they take courses that fulfill major groups' course requirements. The table looks at how the probability students take any "Non-Requirement Major Courses" in an academic year varies over whether students complete majors' course requirements that same academic year.

Strikingly, whether students take any "Non-Requirement Major Course" related to a major group does not seem to depend on whether they make progress in any major group. There are two exceptions: students making progress in the Humanities are 10 percentage points more likely to take a "Non-Requirement Major Course" in the Humanities. Students making progress in the Social Sciences are 11 percentage points more likely to take a "Non-Requirement Major Course" in the Social Sciences. This specific pattern will lead me to over-estimate the amount of learning in the Social Sciences and Humanities, as students who make progress in these majors are also taking more "Non-Requirement Major Courses" related to them.

A.1.8 Major Group and General Graduation Requirements

Focusing on how students take courses to complete majors' course requirements, I do not model how students choose to take courses to satisfy other requirements. These other requirements include distributional or general education requirements and the 120 credits needed to graduate. Ignoring distributional and general education requirements restricts the amount of information students can receive, and limits the students' objective to graduating in a major rather than fulfilling the general requirements for graduation.

The University requires students to satisfy different general distributional course requirements, creating inelastic demand for certain groups of courses. I show distributional course requirements drive students' first-year course-taking, but students specialize their course-taking to satisfy major groups' course requirements afterwards. Table A.3 shows that students' first year of course-taking satisfies both general distributional course requirements: 22.5% ($4.6\% + 15.3\% + 2.6\%$) satisfy only general distributional requirements, 19.1% ($8.1\% + 5.7\% + 5.3\%$) only satisfy major groups' course requirements, and 45.4% ($16.9\% + 15.2\% + 13.3\%$) satisfy both. Starting in their second year, students mostly take courses to satisfy major groups' course requirements. The proportion of courses that satisfy only general distributional requirements drops after students' first year to 8.8% ($2.6\% + 5.3\% + 0.9\%$) in the second year. Meanwhile, the proportion of courses that satisfy only major groups' requirements increases to 52% ($20.2\% + 15.3\% + 16.5\%$) in the second year.

A.2 All Sample Selection Steps: Detailed Empirical and Model Concerns

A.2.1 Empirical Concerns

I create a final sample of students with similar pre-college experiences and who face similar financial constraints. While I am unable to measure students' financial information, I use detailed administrative information about students' incoming characteristics and academic choices to address these empirical concerns.

I first drop 30 students who are missing complete ACT and SAT scores in Step 2.

Students with substantial transfer credit from outside institutions already have progress towards general graduation requirements and major group requirements before starting at the University. The instructor and peer qualities are different, grades come from different distributions, and Advanced Placement exams are completely different. I drop students with more than 24 incoming credits from external institutions or AP exams. Sample selection in Table A.9 shows that I drop around 4500 students in Step 3. This forms almost all students who are dropped due to empirical concerns. Compared to the final sample of students, these dropped students are more likely to be Male and Black, and less likely to be Asian. Dropped students score ten percentile ranks higher in Reading and Math, implying these dropped students are better prepared than students who start at the University with less exposure to college-level courses.

The decision to attend the University half-time instead of full-time is likely neither driven by the desire to learn about abilities nor complete majors. It is likely driven by financial incentives, since students taking a course load with fewer than 12 credits each semester face a different half-time tuition payment schedule. Students may also enroll half-time due to other external shocks. I do not observe any financial aid or tuition expense information in the data, and I drop students who ever enrolled half-time in Step 4. Table A.9 shows that around 250 students ever enroll at half-time or less. This leaves me with a group of students with similar college experiences before starting at the University and whose enrollment decisions are less responsive to external financial or other related shocks.

The last sample selection steps are informed by the dynamic course-taking model and assumptions about how students can make progress across major groups.

A.2.2 Model Concerns

There are four model concerns: students who make no progress across any major groups in one year, students who allocate more than three requirement units (that is, make more than 75% combined progress across all major groups) in one year, students who allocate three requirement units into one major in one year, and students who internally transfer in the first year.

In Step 5 of my sample selection, I drop around 300 students who make no progress across all majors group in one year. To make no progress across all major groups, students are taking courses that satisfy distributional or other general graduation requirements. Table A.9 shows that these 300 students are around six and a half times more likely to drop out (53% to 8%) and half as likely to graduate (45% to 74%) compared to students in the final sample. These students are likely not taking courses outside of CALS, as one or two of them internally transfer out of CALS. Curiously, despite these starkly different outcomes, these 300 students only score in slightly lower Math and Reading percentiles.

In Step 6, I drop around 100 students who allocate more than three requirement units across all major groups in one year. Three requirement units corresponds to between eight to ten courses in one year. Dropped students in Step 6 are similar to the dropped students in Step 7, the 200 who allocate three requirements all in one major group in one year. As expected, these students are likely to drop out compared to students in the final sample. Step 6 and 7 students are more likely to graduate from CALS (92% to 74%) and in return practically none of them internally transfer out of CALS. These students actually score in slightly lower Math and Reading percentiles compared to students in the final sample.

Finally, around 42 students internally transfer out of CALS in their first year. These students score in slightly lower Reading and slightly higher Math percentiles compared to those in the final sample. Aside from their preparation, internally transferring in the first year strongly suggests these students were not interested in graduating in the CALS majors.

A.3 Additional Evidence of Sorting Across Majors

Figure A.3 shows students' first year GPA in each major group over their cumulative progress in that major group over time. Students across quartiles tend to make the same amount of progress over all, suggesting that grades do not play a substantial role in students' course-taking. However, these figures are conditional on students taking courses that satisfy each major groups' course requirements. Students could be choosing not to take courses in these major groups for various reasons. A key reason could be that students with lower beliefs about their match quality in these major groups choose not to take courses in them.

Two patterns stand out from the figures. First, students in the bottom quartile of Natural Sciences GPA make the least amount of cumulative progress in Natural Sciences. This suggests students are reacting positively to their Natural Sciences grades. Second, students in the top quartiles of Humanities and Social Sciences GPA make the least amount of cumulative progress in these two majors, suggesting an opposite story to the Natural Sciences.

A.4 Model Solution, Likelihood Function, and Computational Details

A.4.1 Solving the Dynamic Programming Problem

From the separate definitions for immediate and graduation payoffs $\nu(\mathbb{c}|\mathbb{c} \in \mathbb{L})$ and W_{imt}, and from the learning framework, the student's state variables are her cumulative progress $\overline{\mathbb{c}_{it}}$, and GPA $\overline{g_{it}}$. Then her objective function can be re-written from (1.1) as:

$$\begin{aligned} V_{it}(\overline{\mathbb{c}_{it}}, \overline{g_{it}}, b_{i1}) \quad &= \max\Big\{ \max_{\mathbb{c} \in \mathbb{C}} \Big\{ \mathbb{E}\big[\sum_{m \,\in\, \mathbb{c}} \nu(\mathbb{c}|\mathbb{c} \in \mathbb{L}) | \overline{\mathbb{c}_{it}}, \overline{g_{it}}, b_{i1}\big] + \\ &\beta\Big(\textstyle\prod_{m'=1}^{M} 1 - K_{m'}\Big) \mathbb{E}\big[V_{it+1}(\overline{e_{it+1}}, \overline{g_{it+1}}, b_{i1}) | \overline{\mathbb{c}_{it}}, \overline{g_{it}}, b_{i1}, \mathbb{c}\big] + \\ &\beta\Big(\textstyle\prod_{m'=1}^{M} K_{m'}\Big) \mathbb{E}[W_m(\overline{g_{imt+1}}) | \overline{\mathbb{c}_{it}}, \overline{g_{it}}, b_{i1}] \Big\}, \\ &u^{Bus.}, u^{Oth.}, 0\Big\} \end{aligned} \tag{A.1}$$

To solve this Bellman equation, the student calculates the future value of making each choice $\mathbb{c} : \mathbb{E}\big[V(\overline{e_{it+1}}, \overline{g_{it+1}}, b_{i1}) | \overline{\mathbb{c}_{it}}, \overline{g_{it}}, b_{i1}, \mathbb{c}\big]$. Using backwards induction, the student takes an expectation over future grades and idiosyncratic errors. When the student considers making different choices, she first considers all possible payoffs from the last period, using them to calculate her expected payoffs from all possible choices in the second to last period. This continues until she has expectations for the continuation value for each period's choice.

The idiosyncratic errors $\epsilon_{i\mathbb{c}t}$ play a crucial role in the backwards induction calculation for value functions. Under the assumption that $\epsilon_{i\mathbb{c}t}$ is distributed generalized extreme value with location and scale parameters 0 and τ_t, with a correlation of $\varphi_{\mathbb{L}}$ among requirement unit allocations, McFadden (1978) shows the expectation of the maximum value of future choices conditional on state variables S_{it} and a given choice $\mathbb{c}$ is:

$$\begin{aligned} \mathbb{E}[V_{it}(S_{it+1}|S_{it}, \mathbb{c})] = \quad & \tau_t \mathbb{E}[\gamma_e + \ln\Big(\big(\sum_{\tilde{l} \in \mathbb{L}} \exp[\frac{V_{it}(S_{it+1}, \tilde{l})}{\tau_t \cdot \varphi_{\mathbb{L}}}] \big)^{\varphi_{\mathbb{L}}} \\ & + \sum_{\tilde{l} \notin \mathbb{L}} \exp[\frac{V_{it}(S_{it+1}, \tilde{l})}{\tau_t}] \Big) | S_{it}, \mathbb{c}] \end{aligned} \tag{A.2}$$

where γ_e is the Euler-Mascheroni constant (≈ 0.5772). This formulation is convenient for calculating the probability that students choose any of the $\mathbb{C}$ choices at any time period, and avoids calculating value functions using simulation methods.

The expectation of the value of future choices in (A.2) increases with τ_t. This has an intuitive explanation: as the variance of errors increases, the student's option value for continuing to future periods increases because she is more likely to receive a higher $\epsilon_{i\mathbb{c}t}$ draw. τ_t also plays a leading role for explaining choices within a time period. Consider the case where τ_t is very high – then it is more likely $\epsilon_{i\mathbb{c}t}$ takes on large positive or negative values, leaving the student's choice up to chance.

Uncertainty over future grades is not trivial. If the student receives a low grade in the Natural Sciences, she may choose not to graduate in the Natural Sciences even though she believes she has a higher Natural Sciences match quality because her Natural Sciences GPA is lower. Grades also shift her match quality beliefs, causing the student to change from her expected trajectory of choices. Even if the student perfectly knew her match qualities and did not learn based on her earned grades g_{it}, she may not make the same choices each time period. Suppose she receives positive immediate payoffs from allocating requirement units to Natural Sciences, but does not intend to graduate in it. She actually wants to graduate in Social Sciences because it has a higher graduation payoff. Then one way for her to maximize her utility is to allocate requirement units in Natural Sciences and Social Sciences, and in the last period, only allocate enough to graduate in the Social Sciences.

A.4.2 Likelihood Function of Observed Choices and Grades

In this dynamic choice model, the estimation outcomes of interest are students' observed choices and grades in each major group. I follow the college major choice literature and model students' choices with a nested logit model (with a nest for requirement unit allocations) and grades with ordinary least squares (Arcidiacono, 2004; James, 2011; Stange, 2012). The assumption that $\epsilon_{i\mathbb{c}t}$ is distributed with an generalized extreme value distribution provides an analytical solution for the probability of students' choices over time (McFadden, 1978). Combining the choice and grade likelihood, I estimate all model parameters using maximum likelihood estimation.

Students choose which major groups to receive grades from, and only receive grades g_{imt} if they make progress in that major group. This selection links the two likelihood functions of observed choices, $\mathbb{c}_i$, and grades, g_{imt}. Let Θ represent the set of parameters exclusively involved in the choice likelihood and let Ξ represent the remaining parameters, $(\sigma_m, \phi_m, \Delta)$, which are in the choice and grade likelihoods. Then the joint likelihood of observed choices and grades is:

$$\begin{aligned} \mathfrak{L}_i(\Theta, \Xi) &= \mathfrak{L}_i(\mathbb{c}_i|\Theta, \Xi, X_i, Z_i)\mathfrak{L}_i(g_{imt}|\Xi, X_i, Z_i) \\ \text{where } \mathfrak{L}_i(\mathbb{c}_i|\Theta, \Xi, X_i, Z_i) &= \prod_{t=1}^{T}\Big(\prod_{\mathbb{c}\in\mathbb{C}} \mathbb{P}(\mathbb{c}(S_{it})|\Theta, \Xi, X_i, Z_i)^{\mathbb{1}\{\mathbb{c}_i=\mathbb{c}\}}\Big), \\ \text{and } \mathfrak{L}_i(g_{it}|\Xi, X_i) &= \prod_{t=1}^{T}\Big(\prod_{m=1}^{M} \mathbb{P}(g_{imt}|\sigma_m, \phi_m, \Delta, X_i)^{\mathbb{1}\{\mathbb{c}_{imt}>0\}}\Big) \end{aligned} \tag{A.3}$$

Note that X_i and Z_i are mutually exclusive. X_i contains students' time-invariant characteristics used to estimate prior beliefs b_{i1}, and Z_i are students' time-invariant characteristics used to calculate the major-specific immediate course-taking payoffs u_{imt}.

To calculate the choice likelihood, I solve the dynamic programming problem for each student in the sample, using backwards induction. The total choice set $\mathbb{C}$ contains $\mathbb{L}$ possible requirement unit allocations, and the options to internally transfer to the Business College, internally transfer to Other College, and drop out from the University. I calculate the expected value of each choice for each possible combination of state variables $\overline{\mathbb{c}_{it}}$ and $\overline{g_{it}}$, for each student, $V_{it}(\mathbb{c}, S_{it}|\Theta, \Xi, X_i)$. From the assumption that $\epsilon_{i\mathbb{c}t}$ follows a generalized extreme value distribution with location and scale parameters 0 and τ_t, and are correlated with $\varphi_{\mathbb{L}}$ among requirement unit allocations, the probability of each choice $\mathbb{c}$ is:

$$\mathbb{P}(\mathbb{c}(S_{it})|\Theta, \Xi, X_i, Z_i) = \begin{cases} \dfrac{\exp[\dfrac{V(\mathbb{c})}{\tau_t}]}{1 + \exp[\dfrac{V(\text{Bus.})}{\tau_t}] + \exp[\dfrac{V(\text{Oth.})}{\tau_t}] + G^{\varphi_{\mathbb{L}}}}, \text{if } \mathbb{c} \in \mathbb{L} \\ \dfrac{G^{\varphi_{\mathbb{L}}-1}\exp[\dfrac{V(\mathbb{c})}{\tau_t \cdot \varphi_{\mathbb{L}}}]}{1 + \exp[\dfrac{V(\text{Bus.})}{\tau_t}] + \exp[\dfrac{V(\text{Oth.})}{\tau_t}] + G^{\varphi_{\mathbb{L}}}}, \text{otherwise} \end{cases} \tag{A.4}$$

where $G = \sum_{c \in \mathbb{L}} \exp[\dfrac{V(c)}{\varphi_{\mathbb{L}} \cdot \tau_t}]$, $V(\mathbb{c}) = V_{it}(\mathbb{c}, S_{it}|\Theta, \Xi, X_i, Z_i)$,

and

$$\mathbb{C} = \begin{cases} \{\text{Dropout}, \mathbb{L}, \text{Transfer – Business}, \text{Transfer – Other}\}, & \text{if } t > 1 \\ \{\text{Dropout}, \mathbb{L}\}, & \text{if } t = 1 \end{cases}$$

The grade likelihood is a function of the observed grades and the student's beliefs in each period. The student's major-group-specific beliefs are a function of learning parameters $(\sigma_m^2, \phi_m, \Delta)$, state variables, $\overline{\mathbb{c}_{it}}, \overline{g_{it}}$, and prior beliefs b_{i1}.

$$\mathbb{P}(g_{it}(S_{it})|\Xi, X_i, \mathbb{c}_{it}, \overline{g_{it}}) = \prod_{m=1}^{M} \quad \frac{1}{\sigma_m\sqrt{2\pi}} \exp[-\frac{(b_{imt}(\sigma_m, \phi_m, \Delta, X_i) - g_{imt})^2}{2\sigma_m^2}] \tag{A.5}$$

where b_{imt} is beliefs in each period, and initialized as $b_{im1} = X_i\phi_m$.

Note that the grade likelihood only applies whenever the student is observed to make progress in that major. This comes from the assumption in the dynamic course-taking model that students only receive grades if they made progress in that major group.

A.4.3 Backwards Induction Computation

The main drawback of looking at how students allocate requirement units instead of declaring majors is the exponentially more state variable combinations to consider. There are six state variables: students' cumulative progress, and GPAs across three major groups.

I discretized how students complete major groups' course requirements into intervals of 25%. Then students can only make choices as long as all their cumulative progress is less than 100%, or four requirement units, within a major group. Cumulative grades are by nature continuous, and I discretize them based on observed variation in students' cumulative grades.

For each major group, I use the 10th, 25th, 75th, and 95th percentiles as possible cumulative grade values over time. I discretize the grades students can receive to be 0,2,3, or 4. I do not include grades of 1 as they occur with very low frequency, as shown in Figure A.2.

Discretizing grades and progress to completing course requirements implicitly interpolates value functions across cumulative grade values. For each allocation of requirement units students make, they have different probabilities of receiving a vector of grades. The requirement unit allocation they make and vector of earned grades determines how students transition from one set of state variables values to another.

I assume that students receive one grade regardless of whether they allocate one or two requirement units. Basic statistics says otherwise: the probability of receiving two "A" grades is lower than the probability of receiving one "A" grade. Note this also affects the probability of transitioning between state variable values, and how cumulative grades are calculated. Students who allocate one requirement unit are more likely to have some cumulative grades than students who allocate two requirement units. Fully accounting for this means keeping track of students' entire sequences of grade and choice histories, rather than cumulative progress and GPA, further increasing computational burden.

Assuming that students can allocate one to three requirement units across major groups, and can allocate at most two requirement units into one major group, the choice set of requirement unit allocations $\mathbb{L}$ has sixteen different combinations. For each combination, students can receive different combinations of grades. The requirement units and corresponding grades result in 232 different ways to change state variables in each year.

I show below the number of state variable combinations that need to be calculated for each student in the sample. In the first year, all students have the same state variable combination (zero for all). Since there are 232 different possible outcomes, there are 232 possible state variable values in the second year. Each student needs to consider $232 \times 232 = 53{,}824$ different future combinations. Many of these combinations are not unique, and several do not have any continuation value because the student graduates. For example, choosing (2,0,0) in the first and second years results in graduating in the Natural Sciences.

Year Student Make Choices	Number of Unique State Variable Combinations	$\times$	Number of Outcomes	$=$	Number of Values for Each Student
1	1		232		232
2	232		232		53,824
3	2,678		232		621,296
4	4,257		232		987,624

Then for each of the unique state variable combinations above, I calculate the expected value of each choice. Each choice-state combination either results in graduation in a major group or continuing at CALS. The distributional assumption of $\epsilon_{i\mathbb{c}t}$, (A.2) gives an analytical solution for the expectation of the maximum of these values. Forward simulations and other methods with simulated draws are also possible (Arcidiacono and Ellickson, 2011).

I use backwards induction to calculate the value functions. Each period students can make a choice, they consider all possible outcomes of grades, resulting cumulative GPAs, and resulting match quality beliefs, from each choice they make. For example, when the student is considering her first choice, she first calculates all possible outcomes from all possible choices in the final year. Using these values, she forms her expectation from each possible outcomes for all possible choices in the fourth year, using (A.2). This continues backwards until she has the expected values from all possible choices in the first year.

Once all the value functions for each student at each possible state variable combination over time are calculated, these values are used to calculate the choice likelihood function in (A.4). To reduce the computational load, I use parallel processing to divide the log-likelihood calculation across multiple processors. To further reduce computational burden, I limit estimation to a random subset of around 4,000 students of the final estimation sample.

Individual variation in the value functions comes from individual student characteristics. Increasing the sample size in this case, without any interpolation between student characteristics, can dramatically increase the computational time.[2] To hasten log-likelihood calculation, I divide calculation across multiple computer processors. Each processor calculates value functions for each state variable combination for an even subset of students. The processors then combine log-likelihood functions to numerically calculate the derivative for the log-likelihood with respect to each parameter.

Another way to implement parallel computing is to spread the state variable combinations across processors. However, this dramatically increases the information that is shared across processors. In dividing the sample across processors, processors only share their combined calculated total log-likelihood for a given set of parameter values. Spreading state variable combinations, processors have to share much more information to calculate the log-likelihood function. The processors share multiple times because backwards induction requires knowing the values for all future state variable combinations.

Using this parallel computing strategy, I use the optimizer provided in SciPy (Jones et al., 2001) for estimation. This optimizer package is not "intelligent" of the parallel structure and does not easily communicate the same vector across processors (Dalcin et al., 2005). This makes it difficult to implement an analytical derivative that precisely aggregates vectors. It is more tractable to rely on communicating scalars across processors, and I use the numerical

[2]I previously tried to interpolate across students' characteristics, but this resulted in a non-smooth likelihood function. I also tried interpolating across state variables as done in Stinebrickner and Stinebrickner (2014b), but there was only a small computational speedup.

derivative for optimization. For this reason, I rely on the limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS-B) Algorithm rather than the standard BFGS Algorithm in SciPy which does not communicate the vectors across processors. At the time of this writing, there does not exist a reliable Python module that communicates derivative vectors across processors, where vectors are distributed over the estimation sample. While it is possible to distribute these derivative calculations across parameters, this defeats the purpose of my parallel strategy across the sample size. Finally, the L-BFGS-B requires the log-likelihood and value functions calculated in double precision.

Table A.1: Observed Course-Taking for CALS Students Over All Enrolled Years and Semesters

Year	Semester	Number of Courses Taken	Number of Attending Students
1	Fall	115623	28093
1	Spring	118240	27693
1	Summer	10212	5439
2	Fall	109533	24945
2	Spring	108153	24352
2	Summer	14017	8170
3	Fall	95381	21569
3	Spring	83720	19808
3	Summer	11169	6226
4	Fall	91486	20532
4	Spring	79219	19032
4	Summer	4836	2430
5	Fall	10495	2644
5	Spring	6163	1631
5	Summer	1018	490
6	Fall	1450	396
6	Spring	1193	327
6	Summer	328	154
7	Fall	593	168
7	Spring	520	147
7	Summer	150	71
8	Fall	295	79
8	Spring	201	53
8	Summer	76	36
9	Fall	132	37
9	Spring	121	32
9	Summer	42	20
10	Fall	89	29
10	Spring	102	30
10	Summer	49	14

NOTES – Years of attendance are normalized to the semester students start course-taking. Students can start in Year 1 Fall, Spring, or Summer semesters.

Table A.2: Whether Any "Non-Requirement" Major Courses are Taken in a Year, Conditional on Completing Course Requirements that Year

Major Group	Satisfied Course Requirements	Proportion Taking a "Non-Requirement" Major Course Related to Major Group: Natural Sciences	Human.	Social Sciences	Count
Nat. Sci.	No	0.405	0.685	0.250	39728
Nat. Sci.	Yes	0.384	0.676	0.299	74650
Human.	No	0.415	0.589	0.290	16838
Human.	Yes	0.387	0.695	0.280	97540
Soc. Sci.	No	0.405	0.659	0.186	21774
Soc. Sci.	Yes	0.388	0.684	0.304	92604

NOTES – This shows the proportion of student-years that take a "Non-Requirement Major Course" that is related but does not satisfy any major groups' course requirements. These proportions are shown for student-years that do and do not make complete course requirements in that major group, shown for the final sample of students.

Table A.3: Across Major Groups – Proportion of Courses that Satisfy Major Groups' Course Requirements and General Distributional Requirements

Natural Sciences Courses

	Among All Courses Taken				Among Courses Related to Natural Sciences			
Year	General Dist. Req. Only	Major Req. Only	Both Req.	None but Related	General Dist. Req. Only	Major Req. Only	Both Req.	None but Related
1	0.046	0.081	0.169	0.018	0.146	0.258	0.538	0.057
2	0.026	0.202	0.059	0.023	0.084	0.652	0.190	0.074
3	0.014	0.189	0.020	0.035	0.054	0.733	0.078	0.136
4+	0.011	0.165	0.010	0.058	0.045	0.676	0.041	0.238

Humanities Courses

	Among All Courses Taken				Among Courses Related to Humanities			
Year	General Dist. Req. Only	Major Req. Only	Both Req.	None but Related	General Dist. Req. Only	Major Req. Only	Both Req.	None but Related
1	0.153	0.057	0.152	0.041	0.380	0.141	0.377	0.102
2	0.053	0.153	0.116	0.046	0.144	0.416	0.315	0.125
3	0.023	0.214	0.066	0.060	0.063	0.590	0.182	0.165
4+	0.015	0.221	0.046	0.099	0.039	0.580	0.121	0.260

Social Sciences Courses

	Among All Courses Taken				Among Courses Related to Social Sciences			
Year	General Dist. Req. Only	Major Req. Only	Both Req.	None but Related	General Dist. Req. Only	Major Req. Only	Both Req.	None but Related
1	0.026	0.053	0.133	0.018	0.113	0.230	0.578	0.078
2	0.009	0.165	0.073	0.023	0.033	0.611	0.270	0.085
3	0.004	0.238	0.042	0.041	0.012	0.732	0.129	0.126
4+	0.006	0.196	0.026	0.071	0.020	0.656	0.087	0.237

NOTES – Shows the proportion of courses that satisfy each major groups' course requirements or general distributional course requirements.

Table A.4: List of Academic Subjects Used to Measure Coursework Related in Major Groups, As Shown in Table A.3

- **Natural Sciences**
  Geological Sciences
  Earth & Environment
  Biology
  Chemistry
  Astronomy
  Math & Systems
  Physics
  Statistics
  Geography

- **Social Sciences**
  Political Science
  Psychology
  Anthropology
  Economics
  Sociology

- **Humanities**
  Spanish
  French
  German
  Italian
  Asian Languages
  Russian & Slavic Languages
  Classicals and Architecture
  Latin
  International Studies
  Women's Studies
  American Culture
  Ethnic Studies
  Philosophy & Linguistics
  Religion and Theology
  Film Studies
  Communications
  English and Literature
  History
  Art History

NOTES – These are subjects that are used to determine if a course is relevant to these major groups even though the course did not count towards satisfying that major group's requirements. This list is **not** used to identify courses that satisfy major groups' course requirements. To keep the anonymity of the University, several of these subject names have either been edited or combined with more common subjects. Five to seven subjects with small student enrollment and unique names are omitted.

Table A.5: Subjects that Neither Count Towards Any Major Groups' Course Requirements, Nor are "Related" – Ten Most Common Overall among CALS Students

		Total Taken Courses Each Year				
Subject	Total	1	2	3	4	After 4th
Afroamerican & African Studies	6504	1556	1553	1542	1646	207
Ensemble	6286	2071	1778	1295	1062	80
Study Abroad	3583	42	362	2443	677	59
Romance Languages	3348	2173	862	175	124	14
Dance	3327	421	684	743	1390	89
Writing Center	3236	2011	290	559	328	48
Architecture ♠	3178	836	1974	240	115	13
Education ♣	3162	432	762	758	1115	95
Scholars Program	2836	2580	150	30	72	4
Movement Science §	2425	809	867	350	370	29
Art Design ♠	2369	508	576	464	716	105
Residential College Humanities	2268	857	632	338	404	37
Climate & Space Sciences ▷	1883	282	515	531	517	38
Theatre & Music ♠	1723	312	498	395	433	85
Musicology ♠	1597	398	474	336	355	34
Sports Management §	1317	412	646	140	108	11
Electrical Engineering ▷	1294	123	225	304	464	178
Engineering ▷	1282	501	403	192	160	26
Film	1090	131	284	377	298	
Physiology §	1027	5	159	447	355	61

NOTES – This shows the number of courses taken in these subjects overall, and over the first four years, and all years afterwards. If the subject is not part of the College of Arts and Liberal Sciences, the College it belongs to is listed.
♠ - Art-Related Colleges, ♣ - College of Education, § - Medical-Related Colleges, ▷ - Engineering College

Table A.6: Correlation Between Taking Any "Non-Requirement Major Courses" related to One Major Group with Taking Any "Non-Requirement Major Course" related to Another Major Group

	Natural Sciences	Humanities	Social Sciences
Natural Sciences	1	-	-
Humanities	0.034	1	-
Social Sciences	-0.014	-0.009	1

NOTES – Correlations between taking "Non-Requirement Major Courses" related to different major groups are calculated using all student-year observations in the final sample.

Table A.7: Number of Students with Truncated Time Due to Inferred Completion Times and Internal Transfers

	Inferred Completion in Major Group			Internal Transfer	
Year	Nat. Sci.	Human.	Soc. Sci.	Business	Other
1	0	0	0	0	50
2	1	12	13	1949	1124
3	244	1130	1551	667	1776
4	0	0	0	0	0
Total	245	1142	1564	2616	2950

NOTES – This is shown for the subset of student whose inferred and actual last period differ, for the sample of students after Step (4) in Table A.9. The inferred completion year refers to the year immediately after the student finishes the requirements in any major group.

Table A.8: Internal Transfer Destinations for CALS Students: Before Model Concerns and Final Sample

	Before Model Concerns Years				Final Sample Years			
Destination	1	2	3	4+	1	2	3	4+
Art Related	17	159	478	70	0	158	476	69
Business	0	1955	685	41	0	1955	684	39
Engineering	0	435	164	32	0	435	164	32
Medical Related	25	541	392	531	0	538	387	519
Public Policy or Education	0	11	788	139	0	8	779	133
Total	42	3101	2507	813	0	3094	2490	792

NOTES – This shows the subset of students in the final sample on Table A.9 who internally transfer to Other Colleges at the University. Names of Colleges are generalized to preserve anonymity.

Table A.9: Sample Selection Table: Empirical and Model-Driven Concerns

	Start	Empirical Concerns			Model Concerns				Final
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Time-Invariant Characteristics:									
Female	0.563	0.645	0.481	0.474	0.618	0.530	0.562	0.478	0.574
	(0.496)	(0.486)	(0.500)	(0.500)	(0.486)	(0.503)	(0.497)	(0.501)	(0.495)
Black	0.056	0.097	0.008	0.033	0.090	0.076	0.151	0.044	0.062
	(0.231)	(0.301)	(0.088)	(0.179)	(0.287)	(0.267)	(0.359)	(0.206)	(0.241)
Asian	0.138	0.290	0.236	0.215	0.067	0.167	0.114	0.206	0.126
	(0.345)	(0.461)	(0.425)	(0.412)	(0.251)	(0.376)	(0.319)	(0.406)	(0.331)
Hisp	0.052	0.032	0.024	0.046	0.053	0.106	0.110	0.022	0.055
	(0.222)	(0.180)	(0.153)	(0.211)	(0.223)	(0.310)	(0.313)	(0.147)	(0.228)
Reading Percentile	83.621	0.000	91.260	80.808	82.752	80.288	77.530	82.640	82.811
	(14.718)	(0.000)	(8.965)	(16.339)	(15.840)	(17.048)	(18.107)	(15.826)	(14.756)
Math Percentile	86.222	6.065	93.630	88.987	80.305	81.924	80.059	88.574	85.447
	(13.810)	(23.476)	(7.191)	(10.641)	(17.329)	(15.437)	(18.149)	(13.765)	(13.868)
AP Credits	7.237	3.129	22.600	6.066	4.450	4.742	3.799	9.044	5.362
	(8.479)	(6.318)	(9.495)	(6.544)	(5.867)	(6.115)	(5.586)	(10.583)	(6.075)
Total Transfer Credits	2.946	4.065	6.901	3.079	2.389	2.318	2.123	2.449	2.458
	(4.683)	(5.397)	(7.935)	(4.098)	(4.044)	(4.103)	(3.359)	(4.430)	(3.835)
International (non US Citizen)	0.056	0.452	0.057	0.116	0.025	0.091	0.046	0.169	0.055
	(0.230)	(0.506)	(0.232)	(0.321)	(0.157)	(0.290)	(0.209)	(0.376)	(0.228)
Self-Reported Maximum Parental Education									
$<$ 100K	0.213	0.194	0.144	0.199	0.267	0.258	0.311	0.221	0.221
	(0.410)	(0.402)	(0.351)	(0.400)	(0.443)	(0.441)	(0.464)	(0.416)	(0.415)
$\geq$ 100K	0.401	0.129	0.426	0.371	0.311	0.379	0.320	0.324	0.400
	(0.490)	(0.341)	(0.495)	(0.484)	(0.463)	(0.489)	(0.467)	(0.470)	(0.490)
N	41331	33	4513	305	328	101	213	49	35789

Continued on next page

Table A.9 – *Continued from previous page*

	Start	Empirical Concerns			Model Concerns				Final
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Missing	0.386	0.677	0.430	0.430	0.422	0.364	0.370	0.456	0.379
	(0.487)	(0.475)	(0.495)	(0.496)	(0.494)	(0.485)	(0.484)	(0.500)	(0.485)
Self-Reported Annual Household Income									
HS Graduate	0.079	0.032	0.050	0.142	0.086	0.227	0.096	0.132	0.082
	(0.270)	(0.180)	(0.219)	(0.350)	(0.281)	(0.422)	(0.295)	(0.340)	(0.274)
College	0.284	0.355	0.199	0.275	0.263	0.212	0.329	0.250	0.295
	(0.451)	(0.486)	(0.399)	(0.447)	(0.441)	(0.412)	(0.471)	(0.435)	(0.456)
Advanced Degree	0.451	0.065	0.538	0.401	0.420	0.364	0.361	0.382	0.442
	(0.498)	(0.250)	(0.499)	(0.491)	(0.494)	(0.485)	(0.481)	(0.488)	(0.497)
Missing	0.186	0.548	0.213	0.182	0.231	0.197	0.215	0.235	0.181
	(0.389)	(0.506)	(0.409)	(0.387)	(0.422)	(0.401)	(0.411)	(0.426)	(0.385)
Pre-College Interest									
Natural Sciences	0.435	0.516	0.528	0.526	0.265	0.530	0.511	0.625	0.423
	(0.496)	(0.508)	(0.499)	(0.500)	(0.442)	(0.503)	(0.501)	(0.486)	(0.494)
Humanities	0.359	0.226	0.379	0.301	0.548	0.364	0.320	0.250	0.356
	(0.480)	(0.425)	(0.485)	(0.460)	(0.498)	(0.485)	(0.467)	(0.435)	(0.479)
Social Sciences	0.357	0.387	0.396	0.278	0.324	0.288	0.251	0.279	0.354
	(0.479)	(0.495)	(0.489)	(0.449)	(0.468)	(0.456)	(0.435)	(0.450)	(0.478)
Other College	0.367	0.161	0.350	0.497	0.319	0.439	0.447	0.390	0.368
	(0.482)	(0.374)	(0.477)	(0.501)	(0.467)	(0.500)	(0.498)	(0.489)	(0.482)
Business College	0.271	0.355	0.282	0.275	0.168	0.258	0.205	0.110	0.272
	(0.444)	(0.486)	(0.450)	(0.447)	(0.374)	(0.441)	(0.405)	(0.314)	(0.445)
Outcomes									
Internal Transfer – Other College	0.099	0.129	0.086	0.781	0.034	0.000	0.000	0.331	0.096
	(0.299)	(0.341)	(0.280)	(0.414)	(0.180)	(0.000)	(0.000)	(0.472)	(0.295)
N	41331	33	4513	305	328	101	213	49	35789

Continued on next page

Table A.9 – *Continued from previous page*

	Start	Empirical Concerns			Model Concerns				Final
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Internal Transfer – Business College	0.084	0.065	0.173	0.099	0.000	0.000	0.000	0.000	0.075
	(0.277)	(0.250)	(0.378)	(0.300)	(0.000)	(0.000)	(0.000)	(0.000)	(0.263)
Drop Out	0.101	0.129	0.093	0.020	0.565	0.000	0.000	0.000	0.098
	(0.302)	(0.341)	(0.290)	(0.140)	(0.496)	(0.000)	(0.000)	(0.000)	(0.297)
Graduation from CALS	0.716	0.677	0.649	0.099	0.401	1.000	1.000	0.669	0.731
	(0.451)	(0.475)	(0.477)	(0.300)	(0.491)	(0.000)	(0.000)	(0.472)	(0.443)
N	41331	33	4513	305	328	101	213	49	35789

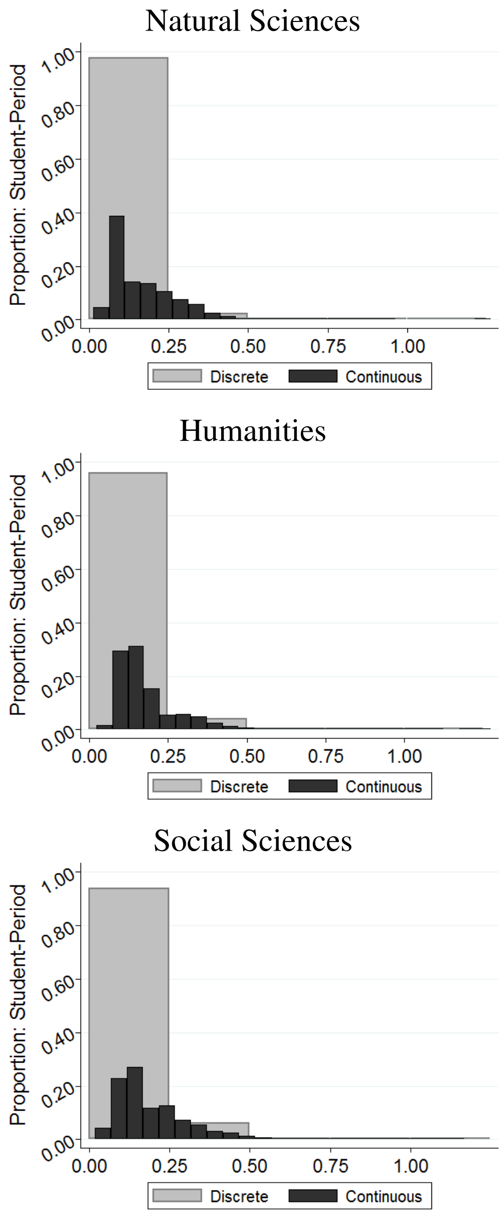
NOTES – Shown Reading and Math percentiles are students' ACT and SAT last reported scores represented as percentiles. If a student has both ACT and SAT scores, then the percentiles are averaged. Transfer credits are all credits from other institutions that are earned before students finish any coursework. Self-reported household income and maximum parental education comes from the Common Application. Advanced degrees include Masters, Doctorates, medical, and law degrees.

Sample Selection Steps:

(1)	Starting Sample
(2)	Missing ACT or SAT Scores
(3)	More than 24 Incoming Transfer or AP Credits
(4)	Ever Enrolled Below Half Time
(5)	Ever Allocated Fewer than One Requirement Unit Total

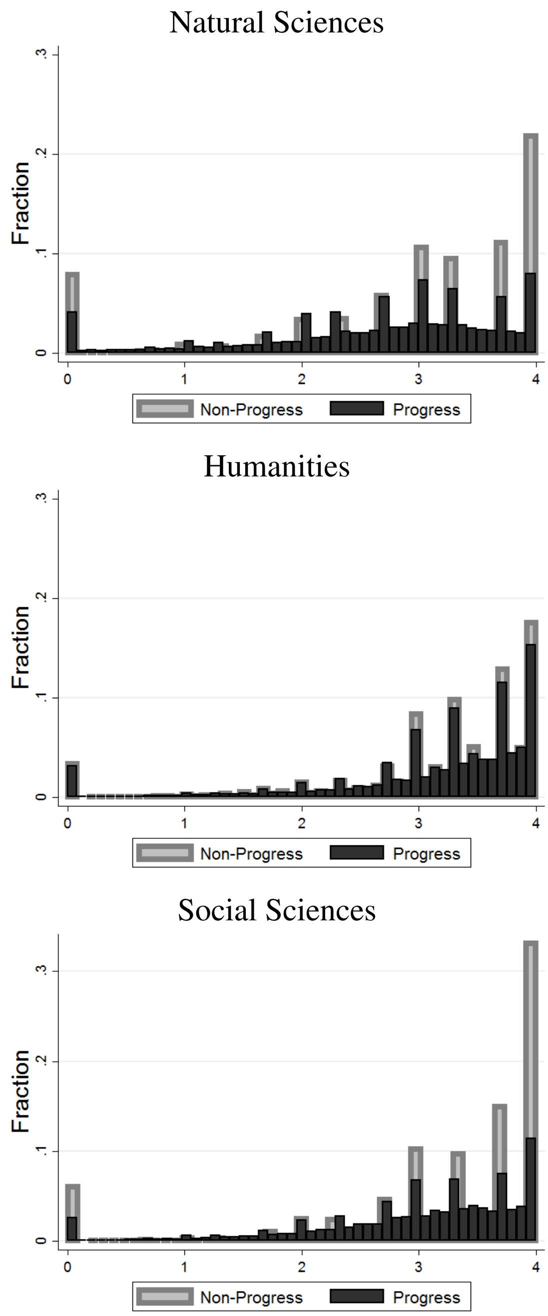
(6)	Overall Choice Set Restrictions
(7)	Individual Choice Set Restrictions
(8)	Low-Share Internal Transfer Restrictions
(9)	Final Sample

Figure A.1: Discretizing Progress Over Time –
Distributions of Discretized and Actual Continuous Progress



NOTES – This is shown on the full starting sample, before dropping any observations due to the sample restrictions shown in Table A.9. Discretized progress is over intervals of 0.25. Continuous progress in red (darker) is coded to be represented with the corresponding blue (lighter) discretized progress.

Figure A.2: Distributions of Grades from Progress and Non-Progress Course-Taking



NOTES – Grades are only shown if the student made progress in that major group, or took a course related to that major group. These grades are credit-weighted at the yearly level. The list of subjects that are related to different major groups can be found in Table A.4.

Figure A.3: Cumulative Progress over First Year GPA

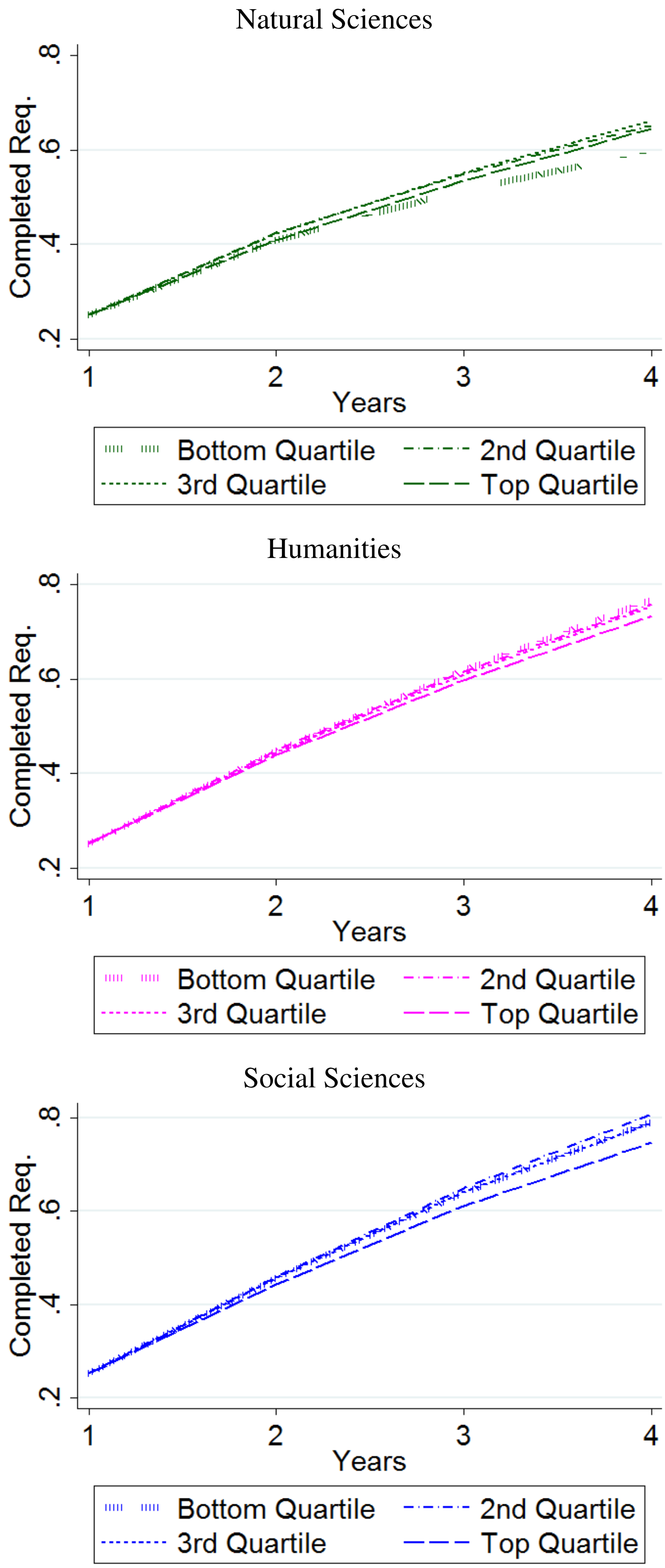
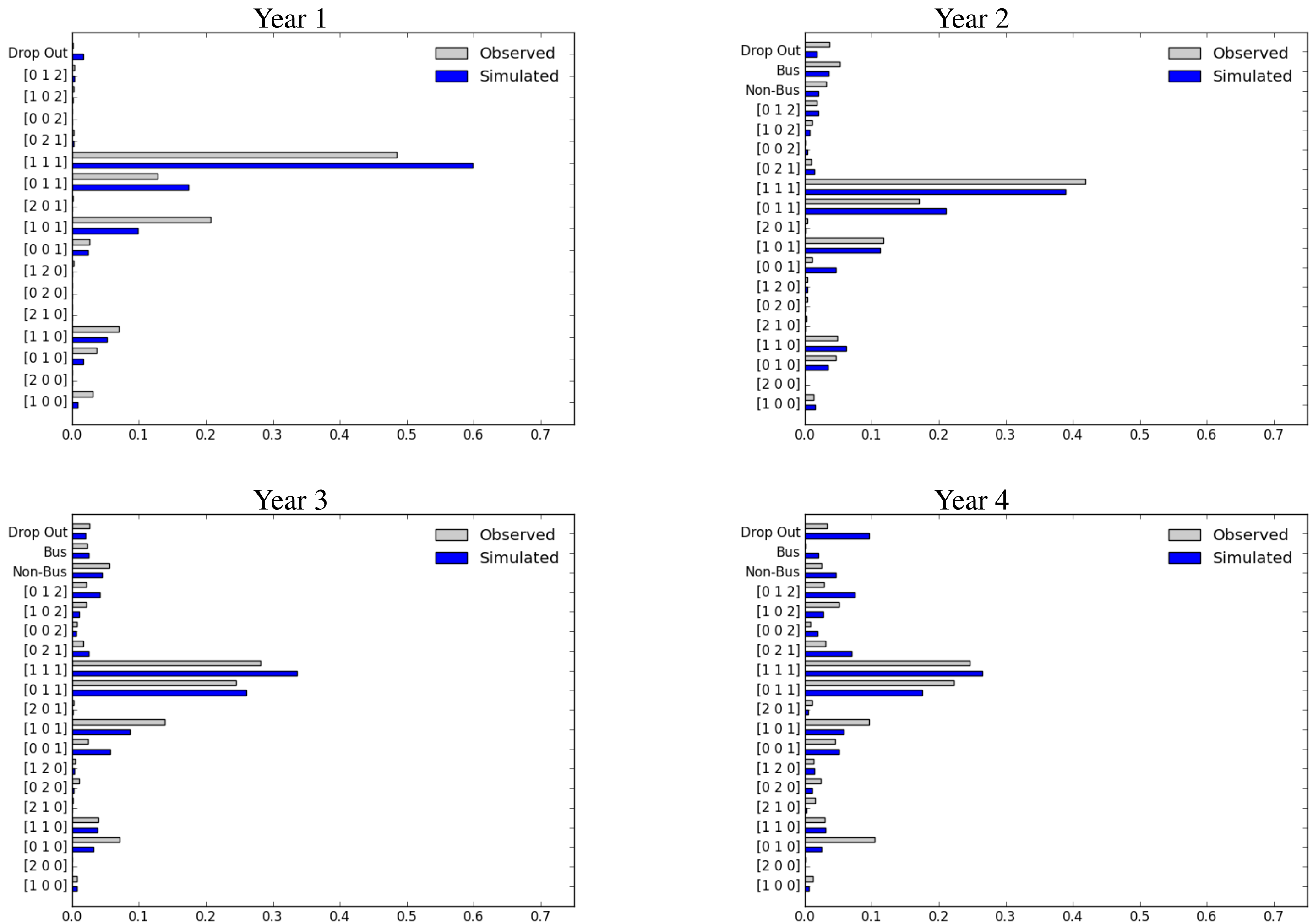


Table A.10: Shares of Requirement Unit Allocations over Time, After Imposing Empirical Considerations on the Sample

Requirement Unit Allocation	Year 1	2	3	4	1	2	3	4	Dropped for Final Sample
0 0 0	23	42	44	162	0.001	0.001	0.002	0.007	
0 0 1	528	268	379	624	0.015	0.008	0.013	0.026	
0 0 2	14	51	139	96	0.000	0.002	0.005	0.004	
0 0 3	1	3	11	52	0.000	0.000	0.000	0.002	X
0 0 4	0	0	0	2	0.000	0.000	0.000	0.000	X
0 1 0	1346	1668	2338	2772	0.037	0.053	0.082	0.117	
0 1 1	4935	6414	8508	5962	0.137	0.203	0.300	0.251	
0 1 2	142	735	761	773	0.004	0.023	0.027	0.033	
0 1 4	0	0	1	0	0.000	0.000	0.000	0.000	X
0 2 0	67	139	301	603	0.002	0.004	0.011	0.025	
0 2 1	499	528	534	872	0.014	0.017	0.019	0.037	
0 2 3	0	0	0	3	0.000	0.000	0.000	0.000	X
0 2 4	0	0	0	1	0.000	0.000	0.000	0.000	X
0 3 0	7	8	11	75	0.000	0.000	0.000	0.003	X
0 4 0	1	0	0	2	0.000	0.000	0.000	0.000	X
1 0 0	675	278	113	182	0.019	0.009	0.004	0.008	
1 0 1	3563	2485	2759	1204	0.099	0.078	0.097	0.051	
1 0 2	48	315	587	1068	0.001	0.010	0.021	0.045	
1 1 0	3129	2204	1497	813	0.087	0.070	0.053	0.034	
1 1 1	20441	15997	10016	7376	0.568	0.505	0.353	0.311	
1 1 3	0	1	0	0	0.000	0.000	0.000	0.000	X
1 2 0	442	223	189	460	0.012	0.007	0.007	0.019	
1 3 1	3	0	0	0	0.000	0.000	0.000	0.000	X
2 0 0	6	3	12	27	0.000	0.000	0.000	0.001	
2 0 1	52	165	65	176	0.001	0.005	0.002	0.007	
2 0 3	0	0	0	2	0.000	0.000	0.000	0.000	X
2 1 0	81	130	76	420	0.002	0.004	0.003	0.018	
2 3 0	0	0	0	1	0.000	0.000	0.000	0.000	X
3 0 0	3	5	1	10	0.000	0.000	0.000	0.000	X
3 0 2	0	0	0	2	0.000	0.000	0.000	0.000	X
3 3 0	0	0	0	1	0.000	0.000	0.000	0.000	X

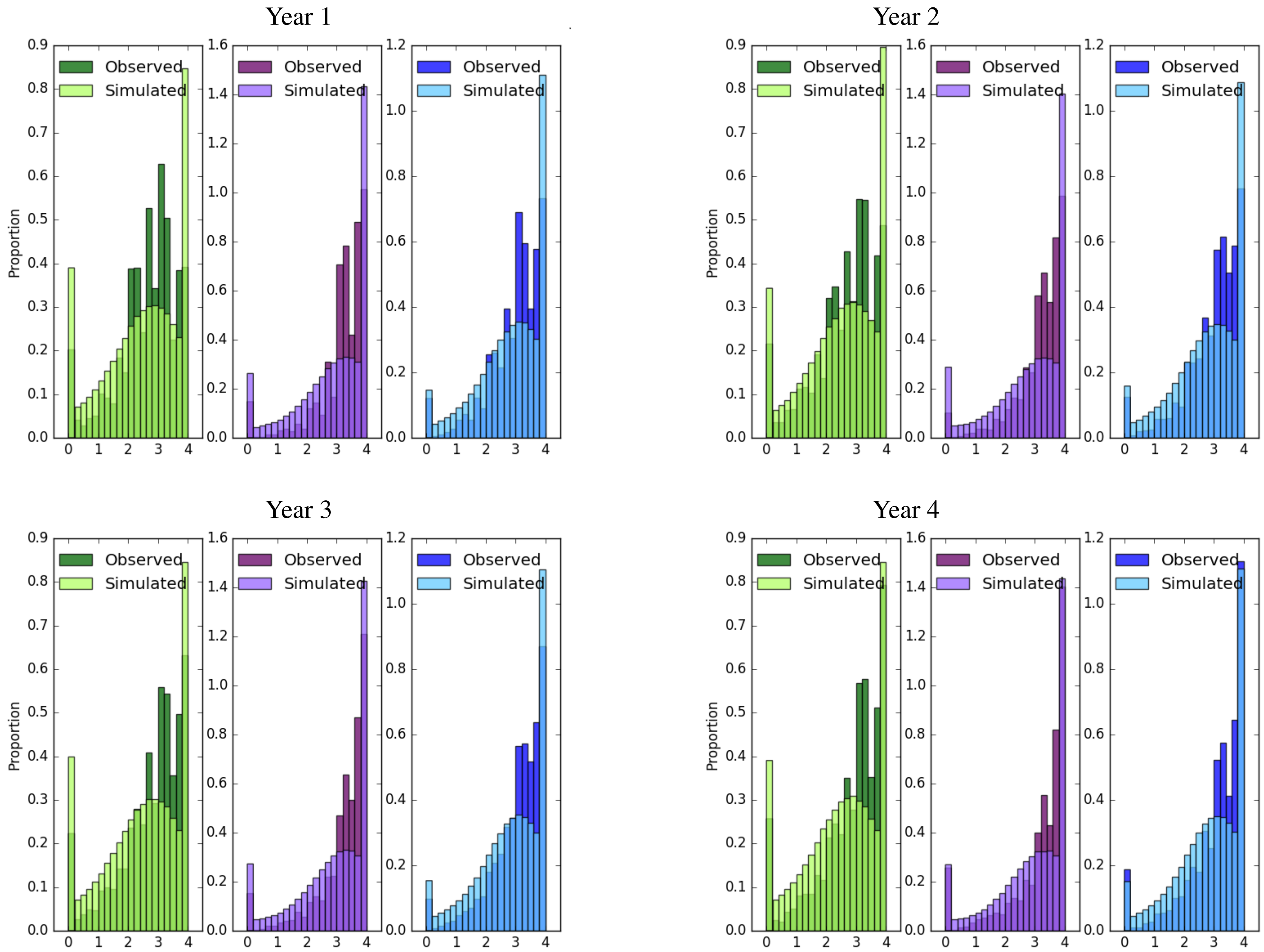
NOTES – Starting from the left, requirement units are allocated in the Natural Sciences, Humanities, and Social Sciences.

Figure A.4: Distribution of Choices: Observed and Simulated



NOTES – Starting from the left, requirement units are allocated in the Natural Sciences, Humanities, and Social Sciences. The denominator in the share is the number of students who are actively making decision in that year. Shares are calculated with 20 simulations over a random subset of 16,000 students not used for estimation.

Figure A.5: Distribution of Earned Grades: Observed and Simulated Grades



NOTES – For each year, observed and simulated grade distributions are shown for (starting from the left) the Natural Sciences, Humanities, and Social Sciences. Grades are conditional on allocating requirement units into a major group. Distributions are calculated with 20 simulations over a random subset of 16,000 students not used for estimation.

Table A.11: Additional Evidence of Model Fit: Student Characteristics Over Aggregate Graduation Outcomes

	Graduate with a Degree At Least in...					
	Nat. Sci.		Human.		Soc. Sci.	
	Obs.	Sim.	Obs.	Sim.	Obs.	Sim.
Reading Percentile	82.812	83.086	84.466	82.633	82.088	82.675
Math Percentile	88.956	86.945	83.184	84.906	85.326	85.282
High School GPA	3.805	3.778	3.733	3.745	3.752	3.749
Black or Hisp.	0.101	0.110	0.124	0.126	0.126	0.123
Black or Hisp. $\times$ Female	0.060	0.069	0.084	0.078	0.083	0.077
Asian	0.187	0.140	0.084	0.120	0.141	0.125
Asian $\times$ Female	0.103	0.077	0.058	0.070	0.082	0.072
Female	0.531	0.560	0.634	0.582	0.590	0.578
International	0.054	0.080	0.057	0.039	0.049	0.057
Reported Household Annual Income						
Less than 100K	0.218	0.232	0.218	0.223	0.217	0.228
More than or equal to 100K	0.407	0.404	0.397	0.383	0.401	0.407
Missing	0.375	0.364	0.385	0.394	0.382	0.365
Maximum Reported Parental Education Level						
High School	0.077	0.081	0.081	0.096	0.082	0.091
College	0.299	0.304	0.297	0.270	0.293	0.294
Advanced Degree	0.444	0.451	0.444	0.445	0.449	0.451
Missing	0.179	0.164	0.178	0.189	0.176	0.164
Prior Interest						
Natural Sciences	0.703	0.674	0.329	0.403	0.438	0.434
Humanities	0.201	0.309	0.480	0.413	0.330	0.366
Social Sciences	0.224	0.320	0.396	0.365	0.377	0.378
Business College	0.140	0.203	0.232	0.243	0.268	0.243
Other College	0.506	0.415	0.311	0.351	0.355	0.357

Continued on the next page

Table A.11 Continued: Additional Evidence of Model Fit: Student Characteristics Over Aggregate Graduation Outcomes

	Internal Transfer to...					
	Other		Business		Drop Out	
	Obs.	Sim.	Obs.	Sim.	Obs.	Sim.
Reading Percentile	81.938	82.826	84.126	82.274	81.010	82.988
Math Percentile	87.166	85.864	92.281	86.689	82.316	85.395
High School GPA	3.800	3.769	3.791	3.746	3.695	3.755
Black or Hisp.	0.080	0.119	0.085	0.113	0.160	0.113
Black or Hisp. × Female	0.044	0.074	0.041	0.062	0.092	0.071
Asian	0.094	0.126	0.198	0.145	0.103	0.118
Asian × Female	0.054	0.073	0.094	0.081	0.061	0.070
Female	0.582	0.581	0.387	0.510	0.597	0.586
International	0.066	0.065	0.060	0.058	0.064	0.056
Reported Household Annual Income						
Less than 100K	0.219	0.208	0.217	0.205	0.229	0.207
More than or equal to 100K	0.399	0.399	0.401	0.396	0.394	0.400
Missing	0.381	0.394	0.381	0.399	0.377	0.393
Maximum Reported Parental Education Level						
High School	0.082	0.065	0.077	0.060	0.086	0.066
College	0.287	0.312	0.296	0.316	0.301	0.308
Advanced Degree	0.457	0.433	0.431	0.433	0.437	0.437
Missing	0.174	0.189	0.197	0.191	0.176	0.188
Prior Interest						
Natural Sciences	0.448	0.468	0.244	0.291	0.386	0.348
Humanities	0.290	0.308	0.303	0.310	0.379	0.334
Social Sciences	0.245	0.289	0.461	0.361	0.364	0.329
Business College	0.185	0.218	0.813	0.648	0.229	0.254
Other College	0.542	0.519	0.210	0.285	0.360	0.350

NOTES – Average of student characteristics gathered from twenty simulations of the baseline model, using a random subset of the final sample that was not used for parameter estimation.

Table A.12: Are Counterfactual Students Different?
Differences between Students Characteristics: Allowed Learning or not from Imposed First Year Course Regime over Aggregate Graduation Outcomes

	Natural Sciences				**Humanities**				**Social Sciences**			
	Baseline Simul.	No Learning	Assigning 111 w/	Assigning 111 w/o	Baseline Simul.	No Learning	Assign 111 w/	Assign 111 w/o	Baseline Simul.	No Learning	Assign 111 w/	Assign 111 w/o
Reading Percentile	83.086	83.499	83.375	83.380	82.633	82.668	82.751	82.659	82.675	82.551	82.565	82.568
Math Percentile	86.945	87.692	87.182	87.111	84.906	84.739	85.061	84.850	85.282	85.105	85.218	85.158
High School GPA	3.778	3.788	3.781	3.780	3.745	3.742	3.748	3.746	3.749	3.746	3.748	3.748
Black or Hisp.	0.110	0.101	0.103	0.103	0.126	0.127	0.124	0.125	0.123	0.126	0.125	0.125
Black or Hisp. × Female	0.069	0.062	0.064	0.064	0.078	0.080	0.078	0.078	0.077	0.079	0.077	0.078
Asian	0.140	0.141	0.138	0.136	0.120	0.118	0.120	0.121	0.125	0.125	0.126	0.125
Asian × Female	0.077	0.077	0.077	0.075	0.070	0.070	0.070	0.071	0.072	0.071	0.072	0.072
Female	0.560	0.555	0.562	0.561	0.582	0.589	0.580	0.586	0.578	0.581	0.576	0.579
International	0.080	0.077	0.071	0.071	0.039	0.036	0.043	0.042	0.057	0.058	0.058	0.058
Reported Household Annual Income												
Less than 100K	0.232	0.232	0.229	0.230	0.223	0.220	0.223	0.222	0.228	0.229	0.228	0.229
More than or equal to 100K	0.404	0.402	0.402	0.404	0.383	0.380	0.385	0.385	0.407	0.412	0.408	0.407
Missing	0.364	0.366	0.369	0.366	0.394	0.400	0.392	0.393	0.365	0.359	0.364	0.364
Maximum Reported Parental Education Level												
High School	0.081	0.083	0.083	0.084	0.096	0.096	0.094	0.093	0.091	0.092	0.090	0.090
College	0.304	0.304	0.300	0.301	0.270	0.262	0.274	0.273	0.294	0.296	0.295	0.296
Advanced Degree	0.451	0.449	0.450	0.448	0.445	0.447	0.444	0.445	0.451	0.452	0.450	0.450
Missing	0.164	0.164	0.167	0.168	0.189	0.195	0.188	0.188	0.164	0.160	0.165	0.164
Prior Interest												
Natural Sciences	0.674	0.656	0.577	0.569	0.403	0.390	0.419	0.411	0.434	0.443	0.448	0.444
Humanities	0.309	0.314	0.330	0.329	0.413	0.423	0.400	0.402	0.366	0.364	0.362	0.360
Social Sciences	0.320	0.321	0.337	0.333	0.365	0.366	0.362	0.363	0.378	0.383	0.372	0.371
Business College	0.203	0.206	0.225	0.222	0.243	0.241	0.244	0.241	0.243	0.235	0.241	0.237
Other College	0.415	0.408	0.389	0.389	0.351	0.347	0.359	0.355	0.357	0.357	0.363	0.362

Table A.12 Continued:

	Business College				Other College				Drop Out			
	Baseline Simul.	No Learning	Assigning 111 w/	Assigning 111 w/o	Baseline Simul.	No Learning	Assign 111 w/	Assign 111 w/o	Baseline Simul.	No Learning	Assign 111 w/	Assign 111 w/o
Reading Percentile	82.826	82.636	82.480	82.499	82.274	82.286	82.163	82.045	82.988	82.581	82.736	82.981
Math Percentile	85.864	85.430	85.248	85.053	86.689	86.647	86.327	86.201	85.395	84.919	85.039	85.764
High School GPA	3.769	3.762	3.757	3.756	3.746	3.744	3.737	3.734	3.755	3.747	3.750	3.757
Black or Hisp.	0.119	0.121	0.126	0.125	0.113	0.112	0.116	0.118	0.113	0.119	0.117	0.117
Black or Hisp. × Female	0.074	0.076	0.080	0.079	0.062	0.061	0.064	0.065	0.071	0.073	0.074	0.073
Asian	0.126	0.126	0.124	0.125	0.145	0.143	0.145	0.138	0.118	0.117	0.116	0.117
Asian × Female	0.073	0.073	0.074	0.074	0.081	0.079	0.080	0.077	0.070	0.071	0.069	0.067
Female	0.581	0.584	0.589	0.587	0.510	0.497	0.510	0.506	0.586	0.591	0.591	0.578
International	0.065	0.064	0.060	0.062	0.058	0.060	0.060	0.056	0.056	0.056	0.057	0.058
Reported Household Annual Income												
Less than 100K	0.208	0.206	0.203	0.202	0.205	0.208	0.205	0.204	0.207	0.206	0.201	0.205
More than or equal to 100K	0.399	0.402	0.407	0.405	0.396	0.400	0.398	0.399	0.400	0.400	0.402	0.400
Missing	0.394	0.392	0.390	0.394	0.399	0.393	0.397	0.396	0.393	0.394	0.398	0.395
Maximum Reported Parental Education Level												
High School	0.065	0.065	0.062	0.063	0.060	0.059	0.060	0.058	0.066	0.064	0.060	0.067
College	0.312	0.313	0.308	0.308	0.316	0.325	0.314	0.318	0.308	0.311	0.321	0.309
Advanced Degree	0.433	0.435	0.443	0.442	0.433	0.430	0.436	0.431	0.437	0.437	0.430	0.437
Missing	0.189	0.188	0.187	0.187	0.191	0.187	0.190	0.193	0.188	0.188	0.189	0.188
Prior Interest												
Natural Sciences	0.468	0.451	0.419	0.406	0.291	0.272	0.247	0.247	0.348	0.313	0.301	0.345
Humanities	0.308	0.302	0.316	0.317	0.310	0.304	0.315	0.314	0.334	0.335	0.318	0.327
Social Sciences	0.289	0.294	0.312	0.308	0.361	0.366	0.368	0.366	0.329	0.333	0.320	0.334
Business College	0.218	0.217	0.228	0.231	0.648	0.676	0.682	0.681	0.254	0.246	0.256	0.255
Other College	0.519	0.530	0.515	0.518	0.285	0.276	0.275	0.276	0.350	0.341	0.347	0.350

NOTES – Student characteristics from twenty simulations of two counterfactual simulations. The first is where students are forced to choose (1,1,1) in the first year, and can learn from the resulting grades. The second is where students do not learn from the resulting grades.

APPENDIX B

Appendix for Chapter Two

B.1 Subjects Related to Each Major

To calculate the “Grade Shocks” for students, I first calculated the major-specific GPAs. Beyond using the explicit courses that make progress towards that major, I included courses with subjects similar to that major.

For example, Calculus I is not required for the Math major at PHEC (only advanced Mathematics courses count towards the Math major), but certainly students will have a better idea of their Math ability after taking Calculus I. Beyond introductory level courses, other majors such as Psychology and Earth Science have more advanced courses that do not count towards the major.

The correspondence for each major is below. The correspondence for each specific major is available upon request.

Table B.1: Course Subject and Major Correspondence

Major Group	Course Subjects
Math & Physical Sciences	Computer Science, Math, Statistics Astronomy, Physics, Math, Statistics
Life & Earth Sciences	Biochemistry, Molecular Biology, Cellular Biology, Developmental Biology, Chemistry, Biomedicine, Biophysics, Atmospheric, Oceanic and Space Sciences Geology Sciences, Environmental Science, Earth Science
Humanities	North African Studies, Hebrew and Judaic Studies, Yiddish, Linguistics, Women’s Studies, Latin American and Carribbean Studies, Film Studies, International Relations, Asian Studies, Classic Civilization, Classic Languages, Classic Architecture, History, Art History, Ancient Civilization and Biblical Studies, Classical Greek, Hebrew Judaic Studies, French, German, Russian, Slavic Culture, Spanish, Polish, Italian, Modern Greek, Communications, English, Comparative Literature
Social Sciences	Sociology, Political Science, Anthropology, Information Science, Complex Systems
Economics	Economics
Psychology	Psychology, Bioanthropology, Organizational Behavior
Art	Theatre Studies, Creative Writing, Drama, Music Studies

B.2 Estimating Students' Initial Major Ability Beliefs

I use students' estimated initial beliefs in order to calculate the signals students receive from their grades. I used ordinary least squares to estimate students' prior beliefs about their grades. The course-taking model assumes grades are a noisy signal of students' true major-specific abilities. A regression on students' first semester grades minimizes selection bias, which would upwardly bias estimates based on selection into taking first semester courses. For each major, I ran a regression of students' first semester grades in that major as a function of:

- Female
- Black
- Hispanic
- Asian
- ACT/SAT Math and Reading Percentiles
- ACT/SAT Math and Reading Percentiles Squared
- ACT/SAT Math and Reading Percentiles Cubed
- Median Reported Household Income in Students' Recorded Zipcode in 2012
- Median Reported Household Income in Students' Recorded Zipcode in 2012 Squared
- Median Reported Household Income in Students' Recorded Zipcode in 2012 Cubed

If a student did not take a class in these majors during their first semester, I do not include them in this regression. I use the estimated coefficients from these regressions to estimate students' initial major ability beliefs.

Table B.1: Proportion of Student-Semesters for Values of Major Progress

	Bins of Semester Major Progress			
Major Group	0%	(0% - 10%]	(10% - 20%]	(20% - 100%]
Math & Physical Sciences	0.72	0.19	0.07	0.01
Life & Earth Sciences	0.56	0.27	0.15	0.02
Humanities	0.44	0.13	0.36	0.11
Social Sciences	0.58	0.12	0.25	0.06
Economics	0.86	0.02	0.11	0.02
Psychology	0.55	0.19	0.21	0.07
Arts	0.96	0.01	0.02	0.01

Table B.2: Extensive Margin of Making Progress in A Major – Average Marginal Effects from Logit Regressions

	Whether the Student Makes Any Progress in Major						
	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Average Marginal Effects of Calculated Shocks and Own Withdraws							
Own Shock	0.022***	0.009***	0.016***	0.016***	0.036***	0.028***	0.013**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.005]
Own Withdraws	–0.019***	–0.016***	–0.056***	–0.054***	–0.019***	–0.058***	–0.007***
	[0.002]	[0.002]	[0.003]	[0.003]	[0.002]	[0.003]	[0.001]
Math & Phys. Shock		0.005	–0.019***	–0.013***	–0.005*	–0.013***	0.000
		[0.003]	[0.003]	[0.003]	[0.002]	[0.003]	[0.001]
Life & Earth Sci. Shock	–0.019***		–0.011***	–0.010***	–0.018***	–0.026***	0.002*
	[0.002]		[0.003]	[0.003]	[0.002]	[0.003]	[0.001]
Humanities Shock	0.002	0.010***		–0.002	0.000	–0.012***	–0.005***
	[0.002]	[0.002]		[0.003]	[0.002]	[0.003]	[0.001]
Social Sci. Shock	–0.009***	–0.011***	0.003		–0.003	–0.007*	0.000
	[0.002]	[0.002]	[0.003]		[0.002]	[0.003]	[0.001]
Economics Shock	–0.010**	–0.014**	–0.019***	–0.015**		0.010*	–0.003
	[0.004]	[0.004]	[0.005]	[0.005]		[0.005]	[0.002]
Psych. Shock	0.011***	0.016***	–0.012***	–0.007*	0.009***		–0.001
	[0.002]	[0.002]	[0.003]	[0.003]	[0.002]		[0.001]
Arts Shock	–0.012	–0.016	–0.019	–0.013	0.004	–0.017	
	[0.017]	[0.018]	[0.019]	[0.019]	[0.014]	[0.019]	

Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Average Marginal Effects of Previous Cumulative Progress							
Own Cumul. Prog	0.794***	0.897***	0.774***	0.801***	0.778***	0.819***	0.245***
	[0.007]	[0.005]	[0.005]	[0.004]	[0.005]	[0.004]	[0.004]
Math & Phys. Cumul. Prog		−0.183***	−0.098***	−0.005	0.044***	−0.068***	0.002
		[0.008]	[0.007]	[0.008]	[0.005]	[0.006]	[0.003]
Life & Earth Sci. Cumul. Prog	−0.011**		−0.072***	0.147***	−0.028***	0.119***	−0.003
	[0.004]		[0.004]	[0.005]	[0.004]	[0.004]	[0.002]
Humanities Cumul. Prog	−0.116***	−0.132***		−0.114***	−0.067***	−0.077***	−0.008***
	[0.004]	[0.005]		[0.005]	[0.004]	[0.004]	[0.002]
Social Sci. Cumul. Prog	−0.033***	−0.089***	0.014**		−0.024***	0.082***	−0.008***
	[0.005]	[0.005]	[0.005]		[0.004]	[0.005]	[0.002]
Economics Cumul. Prog	−0.021**	−0.138***	−0.079***	−0.141***		0.112***	−0.005
	[0.007]	[0.009]	[0.007]	[0.008]		[0.007]	[0.003]
Psych. Cumul. Prog	−0.042***	0.107***	−0.072***	−0.018***	−0.029***		0.006**
	[0.004]	[0.005]	[0.005]	[0.005]	[0.004]		[0.002]
Arts Cumul. Prog	−0.198***	−0.224***	−0.141***	−0.273***	−0.114***	−0.189***	
	[0.015]	[0.017]	[0.013]	[0.016]	[0.012]	[0.014]	
Average Marginal Effects of Time in-Variant Characteristics							
Black	0.016***	0.013***	−0.008*	0.030***	−0.002	0.028***	0.002

Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
	[0.004]	[0.004]	[0.004]	[0.004]	[0.003]	[0.004]	[0.002]
Asian	–0.005*	–0.005	–0.015***	0.011***	–0.004*	–0.002	0.002
	[0.002]	[0.003]	[0.003]	[0.003]	[0.002]	[0.003]	[0.001]
Hisp	0.009**	0.004	0.004	0.023***	0.003	0.007*	0.001
	[0.003]	[0.004]	[0.004]	[0.004]	[0.003]	[0.004]	[0.002]
Reading Percentile	0.000**	–0.001***	0.000***	0.000*	0.000***	–0.001***	0.000*
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Math Percentile	0.001***	0.000**	–0.001***	–0.001***	0.001***	0.000**	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
2012 Median Zipcode Income	0.002	0.000	0.005**	0.003	0.000	0.000	0.000
	[0.001]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
Prior Interest	0.013***	0.055***	0.025***	–0.008***	0.010***	0.006*	0.016***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.003]	[0.001]
N	265999	254083	254164	265786	268589	264808	269258

NOTES – $*** = p < 0.01$; $*** = p < 0.05$, $* = p < 0.10$

The calculated shocks to these majors are the difference of estimated initial ability beliefs from cumulative major GPAs. Initial priors are estimated using a flexible polynomial of pre-collegiate time-invariant characteristics. A complete list of these covariates can be found in the Appendix. Standard errors are clustered on the student level and reported in brackets. Entry term and semester fixed effects are not reported. In addition to the sample restriction in Table 2.3, each regression is restricted to student-semester observations where students' cumulative progress is less than 100%.

Table B.3: Extensive Margin of Making Progress in A Major – OLS Estimates

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Own Shock	0.027***	0.005	0.043***	0.037***	0.081***	0.066***	0.042**
	[0.005]	[0.005]	[0.006]	[0.005]	[0.005]	[0.005]	[0.016]
Own Cumul. Prog	0.935***	1.016***	0.758***	0.784***	1.179***	0.813***	1.009***
	[0.007]	[0.005]	[0.005]	[0.005]	[0.006]	[0.004]	[0.012]
Own Shock × Cumul. Prog	–0.009	–0.010	–0.083***	–0.065***	–0.150***	–0.108***	–0.043
	[0.014]	[0.008]	[0.012]	[0.011]	[0.014]	[0.011]	[0.060]
Own Withdraws	–0.018***	–0.016***	–0.056***	–0.049***	–0.017***	–0.052***	–0.005***
	[0.002]	[0.002]	[0.003]	[0.002]	[0.001]	[0.002]	[0.001]
Math & Phys Sci. Shock		0.011*	–0.008	–0.011*	–0.004	–0.010	0.001
		[0.005]	[0.005]	[0.005]	[0.004]	[0.005]	[0.002]
Math & Phys Sci. Cumul Prog.		–0.243***	–0.095***	0.002	–0.001	–0.060***	–0.007***
		[0.007]	[0.006]	[0.006]	[0.005]	[0.006]	[0.002]
Math & Phys Sci. Shock × Cumul. Prog		–0.016	–0.054***	–0.016	–0.001	–0.017	–0.006
		[0.014]	[0.012]	[0.013]	[0.010]	[0.011]	[0.004]
Life & Earth Sci. Shock	–0.013**		–0.022***	–0.025***	–0.027***	–0.043***	0.001
	[0.004]		[0.005]	[0.005]	[0.003]	[0.005]	[0.002]
Life & Earth Sci. Cumul. Prog	–0.057***		–0.077***	0.140***	0.002	0.108***	0.001
	[0.004]		[0.004]	[0.004]	[0.003]	[0.004]	[0.001]
Life & Earth Sci. Shock × Cumul. Prog	–0.023**		0.031***	0.039***	0.017***	0.046***	0.002
	[0.007]		[0.008]	[0.008]	[0.005]	[0.008]	[0.002]

Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
Humanities Shock	0.009*	0.023***		–0.004	–0.006	–0.012*	0.000
	[0.004]	[0.005]		[0.005]	[0.003]	[0.005]	[0.002]
Humanities Cumul. Prog	–0.069***	–0.101***		–0.093***	–0.029***	–0.061***	–0.008***
	[0.003]	[0.004]		[0.004]	[0.003]	[0.004]	[0.002]
Humanities Shock × Cumul. Prog.	–0.011	–0.024**		0.004	0.014*	0.001	–0.011**
	[0.007]	[0.008]		[0.010]	[0.006]	[0.009]	[0.004]
Social Sci. Shock	–0.016***	–0.024***	0.017***		–0.013***	0.002	–0.001
	[0.005]	[0.005]	[0.005]		[0.004]	[0.005]	[0.002]
Social Sci. Cumul Prog.	–0.038***	–0.076***	0.017***		–0.039***	0.062***	–0.006***
	[0.003]	[0.004]	[0.005]		[0.003]	[0.004]	[0.001]
Social Sci. Shock × Cumul. Prog	0.012	0.021	–0.043***		0.014	–0.030**	0.002
	[0.010]	[0.011]	[0.012]		[0.007]	[0.012]	[0.004]
Economics Shock	–0.013*	–0.012*	–0.019**	–0.016*		0.019**	–0.007**
	[0.006]	[0.006]	[0.006]	[0.006]		[0.007]	[0.002]
Economics Cumul. Prog.	–0.096***	–0.137***	–0.075***	–0.117***		0.095***	–0.004
	[0.006]	[0.006]	[0.007]	[0.007]		[0.007]	[0.002]
Economics Shock × Cumul. Prog	0.042**	–0.004	–0.005	0.016		–0.072***	0.020***
	[0.015]	[0.014]	[0.017]	[0.016]		[0.017]	[0.006]
Psych. Shock	0.023***	0.037***	–0.003	–0.001	0.020***		–0.002
	[0.004]	[0.004]	[0.005]	[0.005]	[0.003]		[0.002]
Psych. Cumul. Prog.	–0.066***	0.045***	–0.068***	–0.034***	–0.035***		0.000

Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
	[0.003]	[0.005]	[0.005]	[0.004]	[0.002]		[0.002]
Psych. Shock × Cumul. Prog.	–0.024**	–0.038***	–0.028*	–0.011	–0.024***		0.003
	[0.009]	[0.011]	[0.011]	[0.011]	[0.007]		[0.004]
Arts Shock	0.004	0.010	–0.021	0.004	0.018	0.002	
	[0.014]	[0.016]	[0.020]	[0.017]	[0.011]	[0.017]	
Arts Cumul. Prog.	–0.120***	–0.133***	–0.130***	–0.196***	–0.057***	–0.125***	
	[0.007]	[0.009]	[0.013]	[0.010]	[0.006]	[0.009]	
Arts Shock × Cumul. Prog.	–0.005	0.020	0.078	0.037	–0.026	0.103	
	[0.044]	[0.050]	[0.077]	[0.056]	[0.030]	[0.055]	
Female	–0.032***	–0.009***	0.009***	0.006**	–0.023***	–0.005**	–0.004***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]	[0.002]	[0.001]
Black	0.015***	0.015***	–0.008*	0.029***	0.000	0.028***	0.001
	[0.003]	[0.004]	[0.004]	[0.004]	[0.002]	[0.004]	[0.001]
Asian	0.002	0.002	–0.016***	0.014***	–0.002	0.001	0.001
	[0.003]	[0.003]	[0.003]	[0.003]	[0.002]	[0.003]	[0.001]
Hisp	0.008**	0.004	0.005	0.023***	0.001	0.006	0.001
	[0.003]	[0.004]	[0.004]	[0.004]	[0.003]	[0.004]	[0.001]
Reading Percentile	0.000***	–0.001***	0.000***	0.000*	0.000***	–0.001***	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Math Percentile	0.001***	0.000***	–0.001***	–0.001***	0.001***	0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

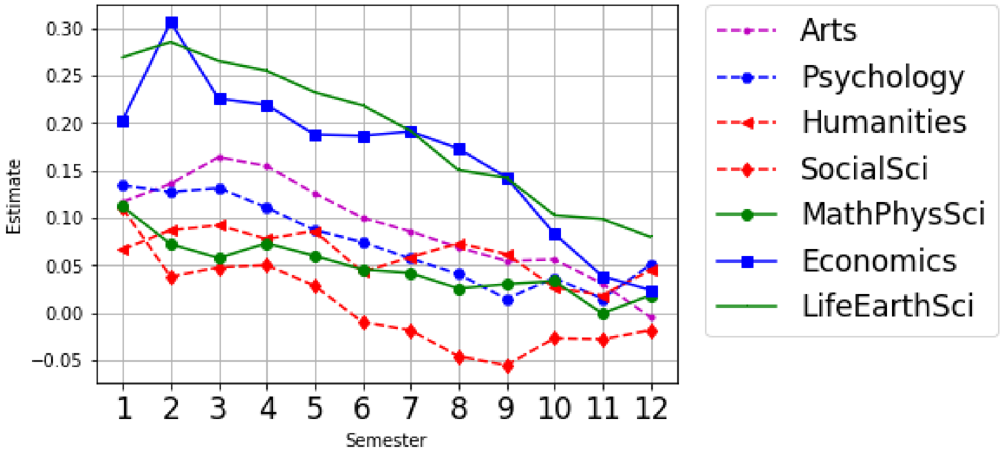
Continues Onto the Next Page...

Continued from Previous Page...

	Math & Physical Sciences	Life & Earth Sciences	Humanities	Social Sciences	Economics	Psychology	Arts
2012 Median Zipcode Income	0.001	0.000	0.005**	0.002	0.000	0.000	0.000
	[0.001]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
Prior Interest	0.023***	0.073***	0.025***	−0.007**	0.045***	0.008**	0.023***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.004]	[0.003]	[0.003]
Constant	0.148	−2.086	0.232	0.367	0.139	0.309	0.068***
	[147.306]	[1425.725]	[48.916]	[47.388]	[52.511]	[41.073]	[0.006]
N	266012	254085	254168	265793	268643	264814	269426
R^2	0.205	0.380	0.155	0.177	0.307	0.166	0.232

NOTES – $*** = p < 0.01$; $*** = p < 0.05$, $* = p < 0.10$
The calculated shocks to these majors are the difference of estimated initial ability beliefs from cumulative major GPAs. Initial priors are estimated using a flexible polynomial of pre-collegiate time-invariant characteristics. A complete list of these covariates can be found in the Appendix. Standard errors are clustered on the student level and reported in brackets. Entry term and semester fixed effects are not reported. In addition to the sample restriction in Table 2.3, each regression is restricted to student-semester observations where students' cumulative progress is less than 100%.

Figure B.1: The Waning Role of Prior Interest in on Making Progress in a Majors' Course Requirements



NOTES – Each point is the estimated coefficient on prior interest in a major on making any progress, separately estimated for each major and students enrolled in that semester. Regressions control for time invariant characteristics: reported gender, ethnicity, zipcode, and Math and Reading percentile scores. Students are not included in a regression if they have have completed a major's course requirements (progress $\geq 100\%$) that semester.

APPENDIX C

Appendix for Chapter Three

C.1 How Recommendations are Calculated

All incoming FAC students are given one of four mutually exclusive recommendations based on their Math SAT or ACT score, High School GPA, and Math Placement Exam. This information is used to calculate a Math Index, and cutoffs in the Math Index determine individual recommendations. While all incoming students are required to take the Math Placement Exam that is conducted by FAC, not all students have SAT, ACT, or High School GPAs.

The first step is to standardize the SAT and ACT scores if the SAT score is not available. If both scores are available, only the SAT Math score is used. If students' Math SAT score is not available, then a Math ACT score is used, where the calculated ACT score is: $20.69 \times$ Math ACT Score$/ + /46.49$. The calculated ACT score is rounded to the nearest hundredths place.

The following formulas are used depending on the availability of these other scores.

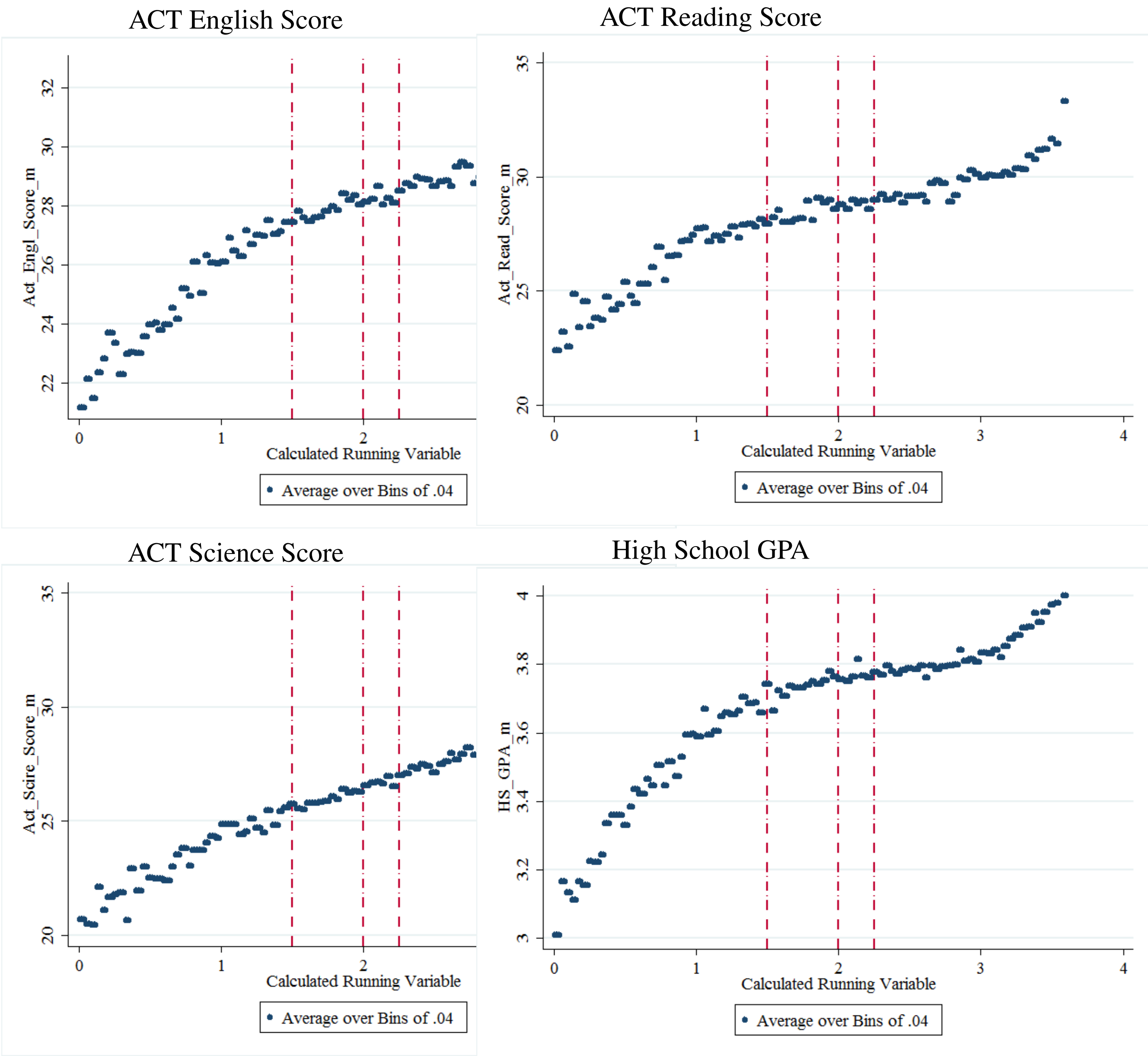
- If no other scores are available:
  Math Index $= -018937 + 0.1454 \times$ Math Placement Score
- If only Math SAT or ACT score is available:
  Math Index $= -1.3272 + 0.11485 \times$ Math Placement Score $+ 0.0023923 \times$ Math SAT
- If only HS GPA is available:
  Math Index $= -2.3981` + 0.12049 \times$ Math Placement Score $+ 0.69086 \times$ HS GPA
- If all information is available:
  Math Index $= -3.4255 + 0.10237 \times$ Math Placement Score $+ 0.0022499 \times$ Math SAT $+ 0.67767 \times$ HS GPA

The recommendations are assigned based on cutoffs in this Math Index:

1. Definitely take Pre-Calculus, if Math Index ≤ 1.5

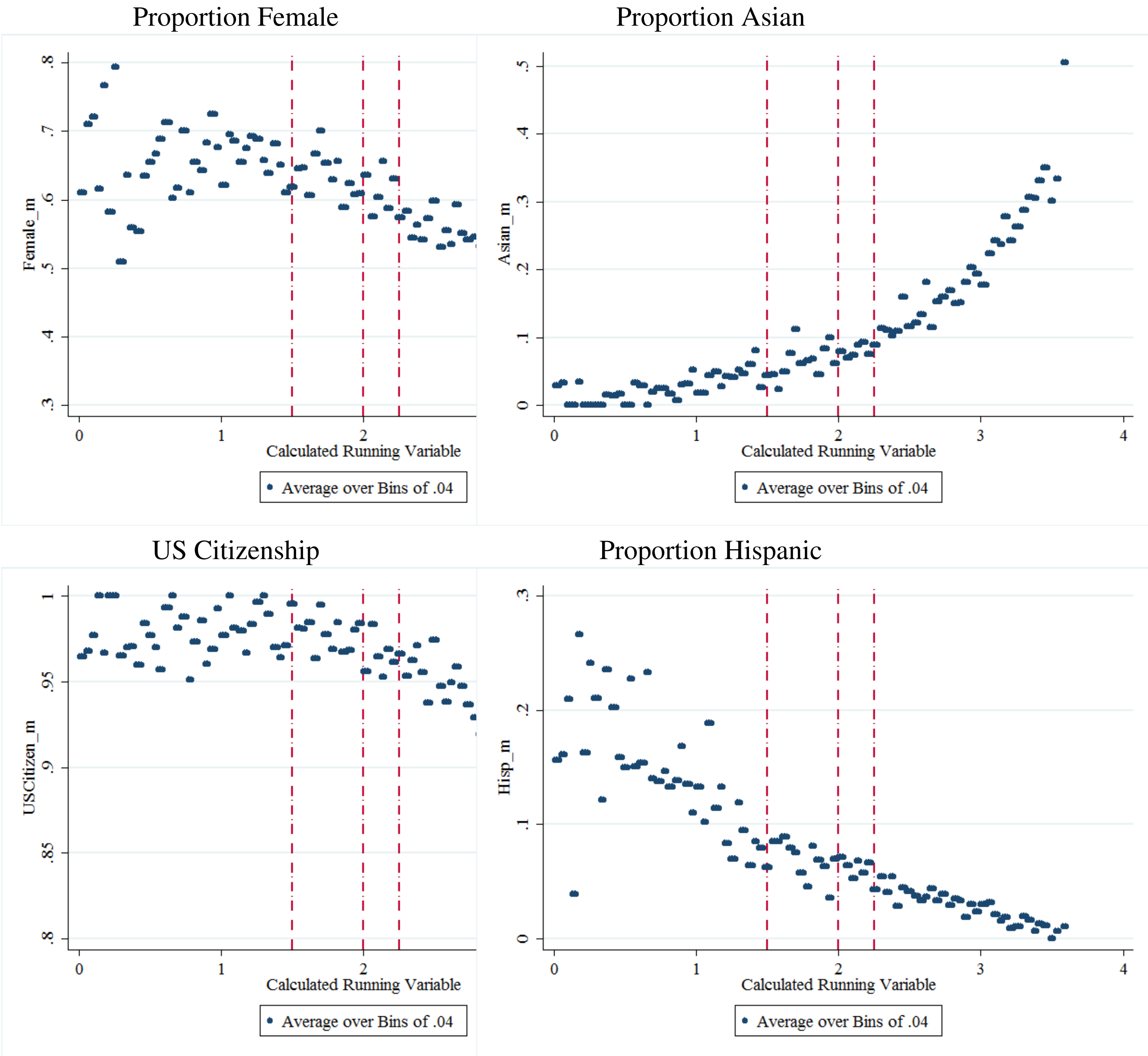
2. Tentatively take Pre-Calculus, if $1.5 <$ Math Index ≤ 2.0
3. Tentative take Calculus, if $2.0 <1$ Math Index < 2.25
4. Definitely take Calculus, if $2.25 <$ Math Index

Figure C.1: ACT English, Reading, and Science Scores, and High School GPA Smoothness Over the Calculated Math Index



NOTES – For each scatter plot, the average of ACT English, Reading, and Science Scores, and High School GPA are shown for narrow intervals of 0.04 of the calculated Math Index.

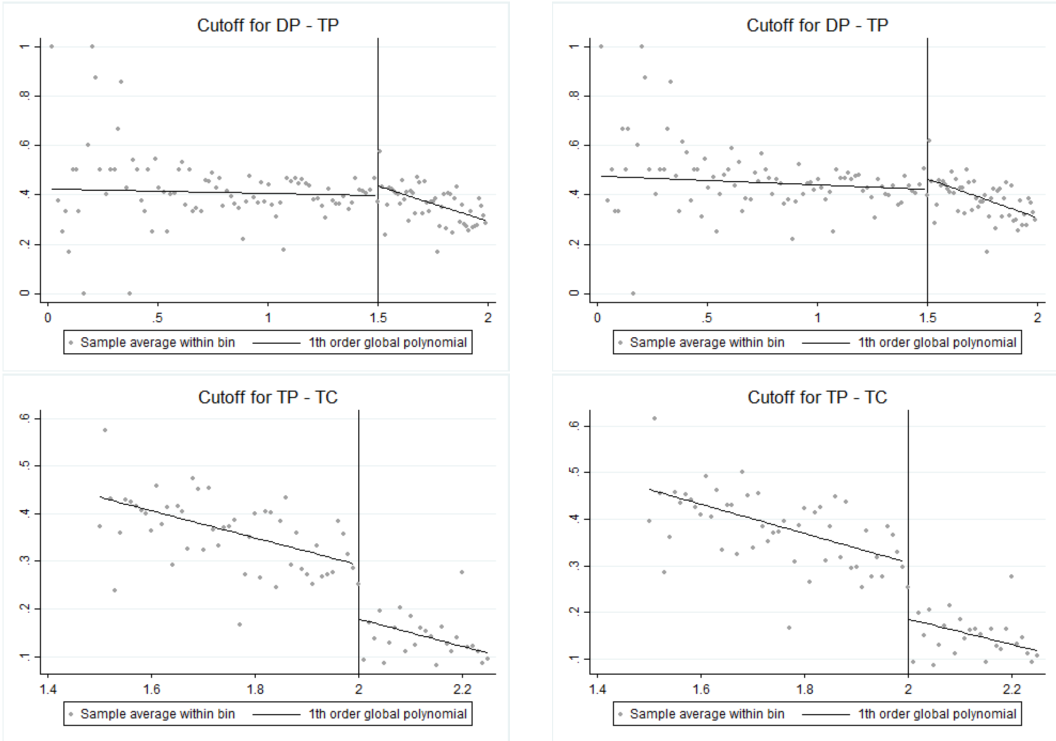
Figure C.2: Proportion of Female, Asian, Hispanic, and US Citizenship Smoothness Over the Calculated Math Index



NOTES – For each scatter plot, the average of Female Students, Asian Students, Hispanic Students, and US Citizens are shown for narrow intervals of 0.04 of the calculated Math Index. US citizenship is recorded upon admission to the FAC.

Figure C.3: Graphical Evidence for Regression Discontinuity for Taking Pre-Calculus at the Three Cutoffs

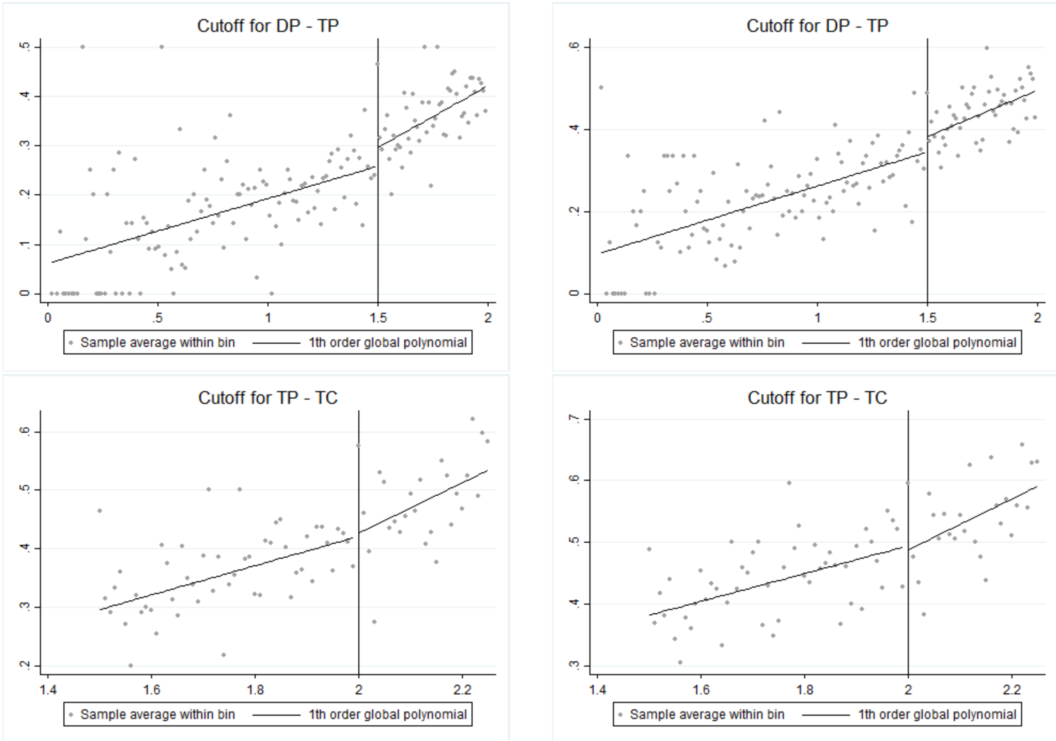
By the Second Semester **By the Eighth Semester**



NOTES – Shown are scatter plot and linear regression lines for the first two cutoffs, between Definite Recommendations to take Pre-Calculus, Tentative Recommendations to take Pre-Calculus, and Tentative Recommendations to take Calculus. Figures for the discontinuity between Tentative and Definite Recommendations to take Calculus are available upon request. Linear regression is done over the entire interval, without choosing an interval around the cutoff.

Figure C.4: Graphical Evidence for Regression Discontinuity for Taking Calculus at the Three Cutoffs

By the Second Semester **By the Eighth Semester**



NOTES – Shown are scatter plot and linear regression lines for the first two cutoffs, between Definite Recommendations to take Pre-Calculus, Tentative Recommendations to take Pre-Calculus, and Tentative Recommendations to take Calculus. Figures for the discontinuity between Tentative and Definite Recommendations to take Calculus are available upon request. Linear regression is done over the entire interval, without choosing an interval around the cutoff.