Work in progress

Math or Science? The Process of Choosing a College Major

Todd Stinebrickner, The University of Western Ontario

Ralph Stinebrickner, Berea College

Abstract: There exist large differences in earnings across college majors, and the number of students that focus on certain specific disciplines, such as those in math and science, is viewed as being an important determinant for the future path of the economy (COSEPUP, 2007). However, while the importance of the college major decision has been well-recognized, much remains unknown about the process by which students arrive at a final major. We take advantage of unique longitudinal data from the Berea Panel Study to provide new evidence about this process, paying particular attention to the choice of whether to major in math and science.

Section I. Introduction

There exist large differences in earnings across college majors, and the number of students that focus on certain specific disciplines, such as those in math and science, is viewed as being an important determinant for the future path of the economy (COSEPUP, 2007). However, while the importance of the college major decision has been well-recognized, much remains unknown about the process by which students arrive at a final major. We take advantage of unique longitudinal data from the Berea Panel Study to provide new evidence about this process, paying particular attention to the choice of whether to major in math and science.

The basic theoretical model motivating past empirical work related to college major has students choosing a major by comparing the expected benefits across the set of possible alternatives (Montmarquette et. al, 2002)). This model highlights the primary difficulty traditionally faced by researchers studying the choice of college major using standard data sources - that beliefs about expected benefits (e.g., expected earnings) are not observed directly for either the major that is chosen by a particular person or for any of the majors that the person considers but does not choose. In response to this difficulty, a small amount of ambitious, recent research has involved collecting survey data specifically for the purpose of studying the choice of college major (Arcidiacono et al. 2010; Zafar, 2008, 2009). Designing survey instruments to draw upon recent methodological advances in the elicitation of beliefs (Dominitz, 1998; Dominitz and Manski, 1996, 1997; Manski, 2004) allows a student's stated major at a point in time to be related directly to his/her beliefs about factors which influence the expected benefits of each major that he/she considers.

While this recent research represents an important step forward, the reality that collecting survey data specifically for a single purpose makes it both difficult to collect longitudinal data and difficult to collect large samples implies that the collected data are not well-suited for the examination of certain central issues.[1] First, while from a theoretical standpoint a student's final major is best viewed as the end result of a process in which he learns about the quality of his match with different majors, there currently exists no evidence about how much uncertainty is present about one's major at the time of college entrance, no evidence about the rate at which uncertainty dissipates over time, and only limited evidence about how uncertainty is resolved.[2] Second, while the importance of understanding decisions to major in specific disciplines, such as those in math or science, are well-recognized, recent research does not provide a primary focus on this issue. To highlight the policy relevance of these two issues, we note that a conclusion that many students seriously consider math/science disciplines but ultimately learn that these disciplines do not represent a good match might have quite different policy implications than a conclusion that very few students enter college with interest in math/science.

Our ability to provide new evidence about the **process** by which students choose a college major, in general, and math/science, in particular, arises because we had the luxury of including questions about college major on our multi-purpose, longitudinal Berea Panel Study (BPS). Of importance for this paper, the BPS, initiated in the fall of 2000, was one of the earliest longitudinal surveys designed to have a specific focus on eliciting expectations information and involved surveying all students in two Berea College cohorts approximately 10-12 times a year during school. Thus, as described in Section II our data include expectations information of relevance for understanding college major at the start of each semester during college, with the first observations coming from the 679 students who participated in our survey at the beginning of their first year.

Blass, Lach, and Manski (2010) describes the importance of allowing agents to express beliefs about a choice in probabilistic form when uncertainty about the choice will be resolved before the final

---

[1] Arcidiacono et al. (2010) uses a single cross-section of 173 students from Duke University across different stages of college. Zafar collects information in the sophomore and junior years at Northwestern University, with 161 students participating in their sophomore year and 117 participating in both years.

[2] Arcidiacono et al. ask respondents whether they have switched their major at any point before the current time. However, students are typically not forced to declare a major until after the first several semesters in college. Then, to the extent that students think about changes to their declared majors when answering the survey question, responses may not capture the amount of uncertainty that exists at early stages of school.

decision is made. Our survey questions used to study college major are unique in this respect. In Section III we find that much individual uncertainty exists at entrance. For example, grouping specific majors into a smaller number of major “groups” (hereafter referred to simply as majors) students assign at time of entrance, an average probability of .43 to the major group that they ultimately end up choosing. After entrance, uncertainty decreases at a roughly constant rate over the first three years. These results are potentially useful for the interpretation of results from other studies which ask students to state their college major at a particular point in college.

Examining trends separately by major, we find that students are quite open to the idea of majoring in math/science at the start of college. Indeed, at entrance: A) the proportion of students who believe that math/science is the most likely major is higher than the proportion for any other major and B) the average perceived probability of choosing math/science is as high as the average probability for any other major. However, by the middle of the third year in college, the proportion of students believing that math/science is the most likely major group decreases by 45 % and the average perceived probability of choosing math/science decreases by 38% so that math/science is ultimately one of the least commonly chosen majors. Important for policy, we find that these changes take place in a non-linear fashion with much of the decrease having occurred by the beginning of the second year even though students are not required to formally choose a major by this time. In contrast to the findings for the math/science major, for all other majors both the proportion and average perceived probability either increase or remain roughly unchanged across semesters. As such, the results in Section III strongly underscore the uniqueness of the math/science field.

In Section IV we examine the determinants of a student’s college major, with a particular interest in understanding the process by which a student decides whether or not to major in math/science. In terms of characteristics that may influence the expected benefits of the various majors, we focus primarily on a student’s beliefs about his academic performance/ability in a particular major and the future income he would receive if he has a particular major. Descriptive statistics suggest the potential importance of these characteristics in explaining the patterns seen in Section III. For example, the evidence suggests that, while on average students enter college believing that math/science is the most difficult of the major groups from a grade perspective, this belief is strengthened considerably over time. Further, particularly striking changes in beliefs are observed for students who start school with a belief that math/science is most likely but subsequently "leave" math/science after the start of college. These students begin school with beliefs that look very similar to those who begin school thinking that math/science is most likely and "stay" in math/science, but finish school with beliefs that look very similar to those who begin school thinking that a major other than math/science is most likely. In general, the results related to academic performance/ability suggest a case where students are pushed rather than pulled out of math/science.

Section V involves estimating models which quantify the relationship between the patterns seen in Section III and beliefs about characteristics that may influence the expected benefits of various majors. To the best of our knowledge, Blass et al. (2010) is the only other paper to estimate models that take into account the uncertainty that a person may have about his choice at the time of a survey. Our work advances this line of research by examining a situation where the source of uncertainty about a real-world choice is well-defined and the degree of uncertainty can be characterized from survey data and is allowed to be heterogeneous across people.[3] In general, the results suggest that beliefs about characteristics such as expected grade point average at a point in time play an important role in determining beliefs about college major at that point in time. More specifically, the results show that changes in beliefs about characteristics associated with math/science play an important role in determining the decline in math/science seen in Section III. Thus, the paper adds to a growing literature which recognizes the importance of learning in determining higher education outcomes.

We conclude in Section VI.

---

[3]Blass et al. (2010) consider a model where consumers must decide among hypothetical bundles of electricity services.

Section II. The Berea Panel Study and a simple model guiding data collection and analysis

II.1. An Overview of the Berea Panel Study

Our analysis is possible because of our access to survey data from the Berea Panel Study (BPS) that were collected specifically for the purposes of this paper. Designed and administered by Todd Stinebrickner and Ralph Stinebrickner, the BPS is a multi-purpose longitudinal survey that takes place at Berea College and elicits information of relevance for understanding a wide variety of issues in higher education including those related to drop-out, college major, time-use, social networks, peer effects, and transitions to the labor market. The BPS consists of two cohorts. Baseline surveys were administered to the first cohort (the 2000 cohort) immediately before it began its freshman year in the fall of 2000 and baseline surveys were administered to the second cohort (the 2001 cohort) immediately before it began its freshman year in the fall of 2001. In addition to collecting detailed background information, the baseline surveys were designed to take advantage of recent advances in survey methodology (see, e.g., Barsky et al., 1997; Dominitz, 1998; and Dominitz and Manski, 1996, 1997) in order to collect expectations towards uncertain outcomes and the factors that might influence these outcomes. Substantial follow-up surveys that were administered at the beginning and end of each subsequent semester document how expectations towards uncertain outcomes and the factors that might influence these outcomes have changed. In addition, time-use surveys were administered eight times a year. Thus, students were surveyed between ten and twelve times a year while in school.

II.2 A simple conceptual framework

Our objective of providing evidence about the process by which a person arrives at a college major involves two primary components. The first component involves characterizing how an appropriate dependent variable for college major changes over time on the path to a final major. The importance of our data for this component can be seen by thinking about issues related to the definition of this dependent variable at the time of entrance (t=1) when we first elicit information pertaining to college major. In a typical discrete choice context, the decision at a particular time is observed. If, one wished to attempt to rigidly follow this framework at the time of entrance, it would be necessary to attempt to utilize some measure of a person's "current" major. From a practical standpoint, because students are not forced to declare a major at entrance, a current major could not be observed in administrative data. More importantly, from a conceptual standpoint, because like most other North American colleges it is not necessary to focus on any specific area in the early semester in order to stay on a path towards graduation, the current major concept is not even well-defined at the time of entrance. More specifically, at Berea roughly half of the first year classes are mandatory under the General Studies component of the liberal arts curriculum with students having substantial flexibility with respect to the other elective classes.

In contrast, the notion of a final major is well-defined. This final major is known with certainty (and could be observed in administrative data) starting at a time t* when the school requires a student to finalize his major choice. However, if one wishes to understand the process leading to a final major, it is necessary to collect information about the final major between t=1 and t* when uncertainty about the final major may remain. As such, the issues here have a direct link to the stated choice literature discussed in Blass et. al (2010) who describe the problems that can arise when a student is forced to "state" a choice when uncertainty exists about a decision that will take place in the future. Then, a key feature of our data is that, at the time of entrance, the first column of Survey Question 1 (Appendix) allows a respondent to express uncertainty about his final major by asking him to report the percent chance that he will ultimately end up with a major in each of seven mutually exclusive and collectively exhaustive major groups: Agriculture and Physical Education (Ag), Business (Bus), Education (Ed), Humanities (Hum), Science including Math (Sci), Professional programs (Prof), and Social Science (SS).[4] Further, the first column of Question 1 was repeated at the beginning of every subsequent semester. This allows us to examine how beliefs change over time on the path to a final major.

The second component of providing evidence about the process by which a person arrives at a

[4] This "percent chance" question was answered after students completed classroom training which, among other things, discussed this type of question in non-education contexts.

college major involves attempting to understand why the dependent variable measuring a person's beliefs about his final major change over time. A simple model guides our data collection and our empirical work in Section V.

A student i arrives at school (t=1) uncertain about his college major. At a time t* he must finalize his choice of major, j. For the sake of discussion we think of t* as happening relatively quickly and abstract from issues related to the utility obtained while in college but before t*. Denote as $U^j(X_i,M_{i,j}\ \epsilon_{i,j})$ the lifetime utility starting at t* that the student receives from choosing j, where X_i is a set of observable permanent characteristics of person i, $\epsilon_{i,j}$ represents the effect on U^j of individual factors that are not observed by the econometrician, and $M_{i,j}$ is the set of major-specific schooling or job characteristics/outcomes (e.g., grade outcomes or future income outcomes) for person i that influence the lifetime benefits of choosing option j. Notationally, we define $M_i = \{M_{i,Ag,}M_{i,Bus}, ..., M_{i,SS}\}$. Our primary interest is in understanding the importance of M_i. As discussed later when we describe the $M_{i,j}$'s that we use in practice, a particular $M_{i,j}$ may influence both the utility received while in school and the utility received after leaving school. If one wanted to understand why a particular characteristic mattered or did not matter at its most basic level, it would be necessary to examine whether the impact of the characteristic was different in the schooling and post-schooling periods. Largely because this task is difficult with identifying information coming only from the college major, our objective is more modest - to examine whether and to what extent characteristics matter in determining changes in the dependent variable. As such, we further simplify things by assuming a simple reduced form for the utility function.[5]

(1) $U^j(X_i,M_{i,j}, \epsilon_{i,j}) = \alpha_j X_i + \beta M_{i,j} + \epsilon_{i,j}$.

At time t* the person is forced to make his final choice and would, therefore, report a probability of one for the chosen major. We make the standard assumption that $\epsilon_{i,j}$ is known by the agent but not the

[5] If, as in Arcidiacono et. al (2010), one assumes that some characteristics only influence utility in school and other characteristics only influence utility after school, then it is possible to put a stronger interpretation on individual coefficients. This is seemingly a reasonable approximation for the types of characteristics used here and in Arcidiacono et. Al (2010), although, as discussed later, this is not guaranteed by theory.

econometrician $\forall$j. In the special case where no uncertainty remains about M_i at t*, U^j is known exactly for each j and the student chooses the option with the highest value. In the more general case where "unresolvable" uncertainty remains about M_i at t*, for each j the student integrates the right side of equation (1) over the distribution of the random variable $M^{t*}_{i,j}$ representing this unresolvable uncertainty and chooses the option with the highest expected utility which we denote $E^{t*}U^j$. In the linear case, this integration results in

(2) $E^{t*}U^j(X_i,M_{i,j}, \epsilon_{i,j})= \alpha_j X_i+\beta\ E(M^{t*}_{i,j})+\epsilon_{i,j}$.

Then, if $E(M^{t*}_{i,j})$ is observed directly in survey data for all j, the econometric analysis at t* follows the standard discrete choice framework. Specifically, while the decision is deterministic from the standpoint of the individual, the econometrician computes the probability of each alternative by integrating over the distribution of $\epsilon_{i,j}$. For example, assuming that $\epsilon_{i,j}$ has an Extreme Value distribution yields the standard logit closed form for the probabilities.

(3) logit -add?

At t=1 and subsequent times before t*, the presence of uncertainty about M_i that will be resolved before t* implies that a person will report a probability of each alternative j. The student knows the decision at t* will be deterministic given $E(M^{t*}_i)=\{E(M^{t*}_{i,1}), E(M^{t*}_{i,2}),...\}$. Then the probability of choice j at time t<t*, $Pr^t(j)$, is the probability that the person ends up with $E(M^{t*}_i)$ such that j is optimal. Letting f be the density representing a person's beliefs about $E(M^{t*}_I)$ as of time t

(3) $Pr^t(j)=\int\int 1(E^{t*}U^j()>E^{t*}U^k \text{ for all } k\neq j)\ f(E(M^{t*}_i))\ dE(M^{t*}_i)$.

Then, before t*, our analysis of reported probabilities at the time of entrance requires the computation of probabilities in equation (3). In part because of the difficulty of constructing a survey question which allows a person to differentiate between resolvable uncertainty and the other sources of possible variation in M_i and in part due to survey space constraints, we do not observe f directly. We describe our approach for constructing f in detail later, here simply noting that our approach takes advantage of several unique questions in the BPS survey.

II.3. The sample

Here we study college major choice using data from the first three years of college. We refer to the start of semesters 1, 2, 3, 4, 5, and 6, respectively, as t=1, t=2, t=3, t=4, t=5, t=6, respectively. Thus, t=1 is the time of entrance and t=6 is the beginning of the second semester of the third year. To maximize our sample size, we combine the 2000 and 2001 cohorts. In total 664, 561, 451, 419, 383, and 376 students in the two cohorts provided legitimate information on Question 1 (introduced in Section II.2) at t=1, t=2, t=3, t=4, t=5, and t=6, respectively.[6] Survey participation rates for the BPS were typically between .85 and .95 for those students who remained at Berea, so that the decrease in response rates above are due primarily to the reality that, as discussed in detail in S&S (2008), the overall drop-out rate is approximately .40 at Berea.

Because we are ultimately interested in how the choice of major evolves over time, it is often useful for our purposes to hold the sample composition constant across the six semesters that we examine. Thus, while we sometimes take advantage of the full sample described above, we often include results for the 371 individuals who answered both the baseline survey and the survey at the beginning of the sixth semester.[7] We refer to this as our composition-constant sample.

The BPS survey data are linked to administrative data to obtain information about a variety of observable characteristics, X_i. We focus primarily on a student's sex and his/her score on the American College Test (ACT). The proportion of students that are male is 41.5% in the full sample. The average (std. deviation) score on the ACT math test is 21.81 (4.14) and the average score on the ACT verbal test is 23.38 (4.51) for students in the full sample. As discussed in Stinebrickner and Stinebrickner (2008), college entrance exam scores at Berea are similar to those at the University of Kentucky and the University of Tennessee.

---

[6] Fourteen students provided "illegitimate" responses in the first column of Question 1 in the sense that the sum of the percent chances in Question 1 was more than 110 or less than 90. For sums that were between 90 and 110, but not equal to 100, we adjusted each percent chance proportionally to make the sum be one.

[7] The 371 number differs from the 376 number seen for t=6 above because five people who reported legitimate values at t=6 had illegitimate values in t=1.

Section III. Beliefs about Major

While it has been recognized that the college major outcome is likely to depend on what a person learns about his match with different majors after arriving at school, to the best of our knowledge, little or nothing is known about how much uncertainty exists about college major at any stage of college. Our evidence comes from responses to Question 1 (Appendix) which was introduced in Section II. For the discussion here we divide the percent chance responses in Question 1 by 100 and describe the responses as probabilities.

Section III.1 Uncertainty about major at different stages of college

Survey question 1 (Appendix) was first administered immediately before the start of the first year. Juster (1966) and Manski (1990) reasoned that, when asked to declare the outcome of a future decision when uncertainty will be resolved before the final decision is made, survey respondents will tend to "state" the alternative with the highest probability as of the time of the survey. Hereafter, we follow this literature by referring to the most likely major at time t as the "stated" major at time t, although we note that this is a misnomer in our context since we have constructed the stated major ourselves. Hereafter, we refer to the stated major at the time of entrance (t=1) as the "starting" major and the stated major in our last observed semester (t=6) as the "final" major. We note that the former is a misnomer because a student is not really forced to start in any particular major, although of relevance later, it is reasonable to believe that students may disproportionately choose electives in the first semester/year from their stated major area. The latter is a misnomer to the extent that some students might not determine their final major until the fourth year of school.

If no uncertainty existed about college major at entrance, each student would assign a probability of one to the stated major. Instead, for our composition-constant sample, the t=1 entry in Figure 1A shows that, on average, students assign a probability of approximately .60 to the starting major. Further, the perception at entrance in Figure 1A may not fully represent the actual amount of uncertainty that exists because many students may ultimately choose a major that is different than the one that they believe is most likely at entrance. The t=1 entry in Figure 1B shows that, on average, students believe at t=1 that the probability associated with the final major is only .44 and that, at t=1, only 5% of students assign a probability of one to the final major. Thus, much uncertainty exists about college major at entrance.

The t=2,...,t=6 entries in Figures 1A and 1B indicate the degree to which uncertainty is resolved after entrance. Both figures show uncertainty decreasing by a fairly constant rate over the six semesters. Then of relevance for interpreting results in recent work, the reasonable notion that there exists a specific period when students are thinking most closely about their college major (Zafar, 2009) does not seem like the appropriate interpretation. Even if students are forced to declare a major by a particular time, it appears that the major choice is a process that begins at entrance so that efforts to encourage the choice of particular majors might need to start very early in school.[8]

Section III.2 Major-specific patterns

The uncertainty about major at entrance seen in Section III.1 raises the possibility that there may be a net inflow from or a net outflow to particular majors over time. To examine major-specific patterns, we begin by constructing, for each semester t and each major j, the proportion of students who have stated major j.

For the composition-constant sample, Figure 2A shows these major-specific proportions across the six semesters. At t=1, the proportion of students with a stated major of Science is higher than the proportion for any other major. However, Figure 2A shows a dramatic decrease of -.09 (.202 to .112), or 45%, in the proportion associated with Science between t=1 and t=6. In contrast, the change in the

---

[8]For example, discussing a survey that takes place in the middle of students' sophomore year Zafar (2009) writes "Since Northwestern University requires students to officially declare their majors by the beginning of their junior year, the timing of the initial survey corresponds to the period when students are actively thinking about which major to choose."

The sample in Arcidiacono et al. (2010) combines students from all stages of college. The results here suggest that the an "intended" major (used in that paper to describer someone at an early stage of college) may mean something quite different than a "chosen" major (used in that paper to describe someone at a later stage).

proportions of the other majors between t=1 and t=6 range from a low of only -.016 to a high +.064.[9] In terms of timing, despite the fact that Figures 1A and 1B show quite constant rates over time, Figure 2A shows that much of the change in the Science proportion between t=1 and t=6 takes place very quickly with (.059/.090)%=66% of the decrease occurring by the beginning of the second year (t=3). This strengthens the conclusion in Section III.1 that the major choice is best viewed as a process in which important changes in beliefs occur early.

While the stated choice can be convenient from a descriptive standpoint, a message that can be taken from Manski (2004) is that a decline in the proportion having the stated major of Science does not necessarily imply that students, on average, had misperceptions about the likelihood of choosing Science.[10] Instead, to understand whether misperceptions about Science existed one needs to examine the data in its original probabilistic form. Figure 2B does this by displaying the perceived probability of each major j averaged over all students in the composition-constant sample. Despite the potential for differences between Figure 2A and 2B, in practice the results in the figures are quite similar. The average perceived probability in the sample is as high for Science at t=1 as for any other major. However, on average, the perceived probability of becoming a Science major decreases by .069 (from .181 to .112), or 38%, between t=1 and t=6 while the change in the average probability for the other majors ranging from a low of -.014 to a high of .048.[11] Again, we find that much of the change for Science happens rather quickly with (.04/.069)%=58% of the decrease in the average Science probability between the start of school (t=1) and the middle of the third year (t=6) occurring by the

---

[9]We reject, at all traditional significance levels, the null that there is no change in the proportion with a stated major of Science between t=1 and t=6 with the test having a t-statistic of 3.39. For each $j \neq$Science, we reject at .05 the null that the change in the stated proportion (between t=1 and t=6) for j is the same as the change in the stated proportion (between t=1 and t=6) for Science.

[10]For example, this could be the case if students with Science as the stated major tend to be less certain about their choice.

[11]We reject, at all traditional significance levels, the null that there is no change in the average perceived probability associated with Science between t=1 and t=6 with the test having a t-statistic of 3.519.. For each $j \neq$Science, we reject at .05 the null that the change in the average perceived probability (between t=1 and t=6) for j is the same as the change in the stated proportion (between t=1 and t=6) for Science.

beginning of the second year (t=3).

Figures 2A and 2B show that, relatively speaking, many students start school with a serious interest in the Science area, but that few students ultimately choose a final major in these areas. To delve further into why Science is unique among majors in this respect, we note that the change in Figure 2A for a particular major j between t=1 and t=6 depends on both: 1) the actual proportion of those having a starting major of j who have a final major of j ("stay" in major j) and 2) the proportion of those having a starting major other than j who have a final major of j ("change" to j). With respect to 1), Figure 3A shows that the sample proportion is lower when j is Science than when j is any of the other majors. With respect to 2), Figure 4A shows that the proportion is lower when j is Science than when j is any of the other majors. Thus, we find that starting in Science is close to necessary but far from sufficient for ending up in Science. In contrast, there exist other majors (most notably Humanities and the Professional major) for which the proportions in both 3A and 4A are large.

Comparing Figure 3A to 3B and 4A to 4B reveals more detail about the source of the overoptimism seen in Figure 2B for the Science major. While Figure 3A shows that, in reality, students are substantially more likely to leave Science than other majors, Figure 3B indicates that, at t=1, students who start in Science believe they are as likely to stay in Science as students in any other major. While Figure 4A shows that students are less likely to change into Science than any other major, Figure 4B shows that students believe they are as likely to change into science as they are to change into any other major.[12] Generally, the results in Figures 3 and 4 strengthen our earlier conclusion that Science is unique among the set of majors.

Section IV. Factors that may influence college major

IV.1 Overall sample means

[12] Figure 4B shows the sample average perceived probability of major j at t=1 for all people who do not have j as their stated major.

In terms of the major-specific characteristics $M_{i,j}$ that influence the lifetime utility of i, we focus primarily on a student's academic performance/ability and the student's future income. The latter is presumably a primary determinant of post-college utility, but could also influence utility while in school if students are able to smooth consumption between the schooling and working portions of their lives. The former is likely to play an important role in determining utility while in school – struggling academically may make studying frustrating, may make it difficult to become interested in course material, and may make school stressful due to a concern about failing out of school – but may also influence a student's post-college utility both by being a determinant of future income and by being related to the extent to which a person enjoys his/her work.

We also examine descriptive results from a survey question which directly asks about a student's current interest in a particular subject area. However, this variable is susceptible to an obvious endogeneity problem. A student's current interest in a major j may be affected by how many classes a person has taken in j in the past. This implies that the change in a student's reported interest for j during a period represents, not only what the student learns about his interest in j during the period, but also the causal impact of the extra exposure to j during the period. Then, the change in the student's reported interest may be largely anticipated by the person rather than representing the type of learning that is of particular interest in this paper. As a result of this endogeneity concern, when we estimate the relationship between $M_{i,j}$ and a student's choice of major we do not include the Interest variable. However, later in this Section (IV.1) we discuss the relationship between our performance/ability measures and the interest measure and describe what this relationship implies about the interpretation of estimated effects in Section V. In Section IV.2 we discuss a circumstance under which the descriptive findings related to the Interest measure might be informative.

Academic performance/ability

With respect to academic performance/ability, the most obvious measure of interest for the student is $AGPA_{i,j}$, a constant representing the average grade point average (GPA) that a person would receive in a given future semester if he were to choose major j. Technically speaking, lifetime utility associated with j might depend on not only the average GPA that a person might receive in a future semester but also the transitory portion of GPA in the future semester. However, the simplifying focus on the average can be motivated by the reality that knowing $AGPA_{i,j}$ is close to sufficient for knowing one's cumulative grade point average at the end of college for j since the sum of the transitory portion of grades will tend towards zero with the number of semesters.

Our beliefs about $AGPA_{i,j}$ come from the second column of Question 1 (Appendix) which asks a student about the GPA that he "would expect to receive in a typical semester in the future" if he had major j. If, at t, the only uncertainty about the future semester GPA in j comes from the transitory portion of grades in j, a student would know $AGPA_{i,j}$ exactly after taking expectations with respect to this transitory component. However, Stinebrickner and Stinebrickner (2008) find that students learn much about their average grade performance after arriving at school. If, at t, uncertainty exists about $AGPA_{i,j}$ the person should also take expectations with respect to this uncertainty when responding to the survey question. Then, drawing on the notation from Section II to let $AGPA^t_{i,j}$ represent a person's beliefs at t about $AGPA_{i,j}$ at time t, the second column of survey question 1 can be viewed as eliciting $E(AGPA^t_{i,j})$.

Figure 5A shows the sample average of $E(AGPA^t_{i,j})$ for each j and each t. Most striking is the pattern related to the Science major. While students do begin school (t=1) with a belief that their grades will be lowest in Science, this belief is strengthened substantially over time. Thus, at first glance, changes after entrance in beliefs about performance in Science have at least the potential to explain the negative slopes of the Science lines in Figures 2A and 2B.

Theory does not suggest whether beliefs about grade performance or beliefs about academic ability per se should be more important in determining major choice. Regardless, given the importance of study effort found in S&S (2004) and S&S (2008b), whether $E(AGPA^t_{i,j})$ should be thought of largely as a measure of beliefs about academic ability per se depends to a large extent on what students believe about their study effort in different majors. On one hand, if students tend to believe that they would expend little effort if they were forced to choose certain majors that might not be of particular interest, low values of $E(AGPA^t_{i,j})$ could be reported largely due to low anticipated effort in j rather than due to beliefs that academic ability is low in j. On the other hand, if students believe that receiving good grades is universally important, they may tend to believe that they will study at least as much when they find courses difficult so that differences in $E(AGPA^t_{i,j})$ across j will tend to reflect differences in academic ability across majors.[13]

Which of the two scenarios is most relevant is an empirical question which can be examined because at time t we elicited the expected amount that a person would study per day in a future semester if he had each potential major group j (survey question not shown). Denoting the reports for major group j at time t as $E(ASTUDY^t_{i,j})$,[14] Figure 5B shows evidence that the second scenario above is more relevant as it pertains to the Science major; while for all t the sample average of $E(AGPA^t_{i,j})$ is lowest when j=Sci, for all t the sample average of $E(ASTUDY^t_{i,j})$ is highest when j=Sci.[15]

We can approach the interpretation of $E(AGPA^t_{i,j})$ more formally by considering a measure $ABILITY_{i,j}$ which represents the average GPA that a person would receive in major j if study effort were held constant across majors. Here we choose a study effort level of 3.0 hours per day, which is approximately the sample average at t=1 across all students and all majors. Since the causal relationship between studying and grade performance in each major j is not observed in our data, it is necessary to make an assumption in order to construct $E(ABILITY^t_{i,j})$, the mean of the distribution describing i's

[13] Of course, some majors may tend to have higher grades than other majors for reasons unrelated to study effort or academic ability.

[14] Consistent with our notation for the GPA variable, we think of $ASTUDY_{i,j}$ as a constant measuring the true average amount a person would study in the future, $ASTUDY^t_{i,j}$ as a RV representing a person's beliefs about $ASTUDY_{i,j}$ at time t and $E(ASTUDY^t_{i,j})$ as the mean of this belief distribution.

[15] A test rejects, at all traditional significance levels, the null that there is no change in $E(AGPA^t_{i,j})$ over time. For each major $j \neq$ SCI, HUM,BUS, a test rejects, at all traditional significance levels, the null that the difference between $E(AGPA^t_{i,SCI})$ and $E(AGPA^t_{i,j})$ is the same at t=6 as at t=1. For each major $j \in \{HUM,BUS\}$ a test rejects at significance levels greater than .07 the null that the difference between $E(AGPA^t_{i,SCI})$ and $E(AGPA^t_{i,j})$ is the same at t=6 as at t=1.

beliefs about $ABILITY_{i,j}$ at time t. We assume that the causal effect of studying is homogenous across j and use the estimate of the causal effect of studying from Stinebrickner and Stinebrickner (2008b).[16] Not surprisingly given the earlier discussion, the message from the sample averages for $E(ABILITY^t_{i,j})$ in Figure 5C is similar to the message from Figure 5A. In the sample, the difference between Science and the other majors at entrance is larger in 5C than 5A and the gap between Science and the other majors increases more in 5C than 5A. Thus, our results support the notion that differences in $E(AGPA^t_{i,j})$ tend to largely represent differences that are not attributable to effort. Given the similarities between 5C and 5A and the reality that creating 5C requires assumptions about the causal effect of studying across majors j, we choose to use $E(AGPA^t_{i,j})$, rather than $E(ABILITY^t_{i,j})$, as our primary measure in the remainder of the paper.

Income

With respect to future income, our measure of interest is $AINCOME_{i,j}$ which represents the average income a person would receive at age 28 if he had major j. The mean of the distribution describing a person's beliefs about $AINCOME_{i,j}$, which we denote $E(AINCOME^t_{i,j})$, comes from the third column of Survey Question 1. Figure 5D shows large decreases in the sample average of $E(AINCOME^t_{i,Sci})$ over time. However, unlike what is seen in Figures 5A and 5C, the decreases for the other majors are similar in nature to those observed for Science.

Current Interest

Figure 5E shows sample averages for a student's interest level in different majors j at time t, $INTEREST_{t,i,j}$, as elicited in the last column of Survey question 1. Of particular interest, there is non-trivial interest in Science at entrance, with Science being the median major in terms of sample average interest at t=1. However, by t=6 the sample average interest in Science has fallen below the average interest in all but one other major (the very specific Agriculture major).

---

[16] S&S (2008b) take advantage in variation in study effort created by whether a student's roommate brought a video game to school. We assume that studying an extra hour per day increase a student's grade point average by .30.

The endogeneity concern mentioned earlier arises from the possibility that, if exposure (or continued exposure) to a particular major is necessary for a person to have interest (or maintain interest) in that major, then the decrease in $INTEREST_{t,i,SCI}$ could be caused by students leaving Science rather than vice-versa. It is this theoretical concern that leads us to exclude the INTEREST variable from our specifications in Section V.

In practice we do find a strong relationship in our data between what a person learns about his academic performance/ability in Science and changes in the person's interest in Science.[17] Such a relationship could be a result of the endogeneity of the INTEREST variable; what a person learns about his academic performance/ability could determine whether or not the person chooses Science, which in turn could influences the person's subsequent current interest in Science. However, a person's academic performance/ability in Science may also have a direct causal effect on a person's interest in Science since students who have a difficult time understanding course material may have difficulty fully appreciating the subject matter. We note that excluding the INTEREST variable in Section V in a case where performance causes interest is not problematic given that our goal in Section V is not to differentiate between the possible avenues through which academic performance/ability in Science could influence whether a person chooses Science.[18] The interpretation in Section V would become difficult if one is concerned about reverse causality - that decreasing interest in Science could be causing the decline in beliefs about academic performance in Science. However, to the extent that study effort is the most likely avenue through which a decreasing interest in Science would effect GPA, this possibility can be examined directly. Consistent with our earlier suggestion that course difficulty may be the driving force in how much a person expects to study in major j, we find no evidence that decreases in

---

[17]Regressing [$INTEREST_{6,i,Sci} - INTEREST_{1,i,Sci}$] on [$E(AGPA^{6}_{i,Sci})$-$E(AGPA^{1}_{i,Sci})$] leads to an estimated effect (std. error) of .599 (.092).

[18]That is, we interpret the total effect of $AGPA_{i,j}$ as capturing a variety of things including a person's interest in a major.

$INTEREST_{t,i,Sci}$ are associated with decreases in $E(ASTUDY^t_{i,Sci})$.[19] The interepretation in Section V would also become difficult if $E(AGPA^1_{i,Sci})$ and $INTEREST_{1,i,Sci}$ were correlated because both are influenced by a common background factor (e.g., a parent in a Science occupation). We discuss this matter more in Section V.

IV.2. Heterogeneity in beliefs about factors

Heterogeneity acorss different observable groups

The fact that beliefs about certain characteristics in M_i, for example $AGPA_{i,j}$ and $ABILITY_{i,j}$, become less positive over time suggests that these beliefs may have the potential to help explain the decrease in the sample average seen in Figure 2B for our primary dependent variable of interest - the perceived probability of choosing science. To further investigate this potential we examine whether there exist differences in beliefs for three (mutually exclusive and collectively exhaustive) groups which have very different patterns for this dependent variable: those who started in Science and stayed in scienStay_Science) , those who started in Science but did not stay in Science (Leave_Science), and those who started in a major other than Science (Start_Other).[20] Intuitively speaking, a belief variable such as $E(AGPA^t_{i,Sci})$ will tend to be successful as an explanatory variable if the patterns associated with $E(AGPA^t_{i,Sci})$ for the three groups are similar to the patterns for the dependent variable for the three groups. More specifically, as can be seen in Figure 6, which is obtained by recalculating the Science component of Figure 2B for each of the three groups, what is needed is that: 1) at t=1 the average value of the belief variable should be similar for the Leave_Science and Stay_Science groups and beliefs for these groups should be substantially different than beliefs for the Start_Other group; 2) by t=6, the average value of the belief variable should be similar for the Leave_Science and Start_Other groups and beliefs for these groups should be substantially different than beliefs for the Stay_Science group.

[19] Regressing $[E(ASTUDY^6_{i,Sci})-E(ASTUDY^1_{i,Sci})]$ on $[INTEREST_{6,i,Sci} - INTEREST_{1,i,Sci}]$ leads to an estimated effect (std. error) of .113 (.085).

[20] It would be desirable to separate the Start_Other group into a Stay_Other and Leave_Other group. However, this is not practical due to the very small number of students move into the Science category (Figure 4A).

Figure 7A, which shows the sample average of $E(AGPA^{t}_{i,SCI})$ from Figure 5A disaggregated into the three groups, suggests that $E(AGPA^{t}_{i,SCI})$ may be a particularly promising explanatory variable. At t=1 the Stay_Science group has views about grade performance in Science that are substantially more positive than the Start_Other group and the gap between the two groups remains relatively constant across semesters. However, the sample average $E(AGPA^{t}_{i,SCI})$ for the Leave_Science group changes dramatically over semesters. Students in this group begin college with beliefs that are very similar to the Stay_Science group, but by t=6 have beliefs that are much more similar to the Start_Other group.[21]

The decision of whether to major in Science will depend on, not only beliefs about grade performance in Science, but also beliefs about grade performance in the alternative majors. To construct an analog to Figure 7A which describes the beliefs of the three groups about grade performance in the alternative majors, we simplify by aggregating the alternative majors into a single Non-Science major (NON-SCI). To create $E(AGPA^{t}_{i,NON\text{-}SCI})$ for person i at time t we take a weighted average of $E(AGPA^{t}_{i,j})$ for all $j \neq SCI$, where the weight associated with j is the student's reported probability that he will choose j conditional on $j \neq SCI$.[22] Figure 7B reveals no evidence of the types of patterns observed in Figure 6; the three groups start school with very similar views about $E(AGPA^{t}_{i,NON-SCI})$ and the views of each group remain quite constant across semesters.

Figure 8A shows the sample averages of $E(ASTUDY^{t}_{i,SCI})$ for the three groups. The lines are quite similar, thereby indicating that the differences between the lines in Figure 7A reflect something closer to ability rather than differences in effort. This point is made more formally in Figure 9A where

---

[21]For t=1, we reject the null that the average $E(AGPA^{t}_{i,SCI})$ is the same for Stay_Science and Start_Other (t-statistic 10.071) and reject the null that the average $E(AGPA^{t}_{i,SCI})$ is the same for Leave_Science and Start_Other (t-statistic 11.379). We cannot reject the null hypothesis that $E(AGPA^{t}_{i,SCI})$ is the same for Stay_Science and Leave_Science (t-statistic .607).For t=6, we reject the null that the average $E(AGPA^{t}_{i,SCI})$ is the same for Stay_Science and Start_Other (t-statistic 11.387) and reject the null that the average $E(AGPA^{t}_{i,SCI})$ is the same for Stay_Science and Leave_Science (t-statistic 5.41). A test of the null hypothesis that $E(AGPA^{t}_{i,SCI})$ is the same for Leave_Science and Start_Other has a t-statistic of 1.97.

[22]If the probabilities associated with the alternative majors are all zero for a particular t, we construct the weights using the probabilities from the most recent period in which the probabilities were not all zero. The variables $E(AINCOME^{t}_{i,NON\text{-}SCI})$, $E(ABILITY^{t}_{i,NON\text{-}SCI})$, $E(ASTUDY^{t}_{i,NON\text{-}SCI})$ and $INTEREST_{t,i,NON\_SCI}$ are constructed in the same way.

the sample average of $E(ABILITY^t_{i,SCI})$ has patterns for the three groups that look very similar to the patterns seen in Figure 7A. Consistent with what was seen in Figure 7B, Figures 8B and 9B reveal little difference in $E(ASTUDY^t_{i,NON\text{-}SCI})$ and $E(ABILITY^t_{i,NON\text{-}SCI})$ across the three groups.

Thus, in terms of beliefs about grade performance/ability, differences between the groups exist primarily because of differences in beliefs about Science rather than differences in beliefs about the Non-Science alternatives. At least in terms of what one learns about academic performance/ability, the evidence suggests that students are pushed rather than pulled out of Science.

With respect to $E(AINCOME^t_{i,SCI})$ Figure 10A shows patterns for the three groups are not nearly as distinct as those seen for the performance/ability measures. At t=1, the Stay_Science and Leave Science groups do have significantly more positive views about future income than the Start_Other group.[23] However, at t=6, while the order of the sample averages is what one would expect from a successful explanatory variable, the differences between the three groups is small and not statistically significant. Figure 10B shows little difference between $E(AINCOME^t_{i,NON\text{-}SCI})$ for the groups. Thus, the results suggest that beliefs about income may not play as important a role as beliefs about performance/ability in explaining the major decision.

In terms of a student's interest in different majors, Figures 11A and 11B indicate that the Leave_Science group begins college (t=1) with average values of $INTEREST_{t,i,SCI}$ and $INTEREST_{t,i,NON\text{-}SCI}$ that are very similar to the Stay_Science group, but ends school (t=6) with average values that are similar to those of the Start_Other group. Thus, the INTEREST variables for the three groups follow patterns that are similar to those observed for the dependent variables in Figure 6. However, as discussed at the start of Section IV, an interpretation of these finding requires taking into account that the change in reported interest for a particular major j between two periods represents both

[23] At t=1, the t-statistic from a test of the null that at t=1 average $E(AINCOME^t_{i,SCI})$ is the same for the Stay_Science and the Start_Other group is 2.636 and the t-statistic from a test of the null that at t=1 average $E(AINCOME^t_{i,SCI})$ is the same for the Leave_Science and the Start_Other group is 2.395. At t=6 the null hypothesis that $E(AINCOME^t_{i,SCI})$ is the same for the Stay_Science and Start_Other groups has a t-statistic of 1.83. The null hypothesis that average $E(AINCOME^t_{i,SCI})$ is equal across groups cannot be rejected for any of the other possible pairs.

what the student learns about his interest in j during the period and the causal impact of taking more classes in j during the period.Thus, it is difficult to know whether and to what extent the increase in $\text{INTEREST}_{t,i,\text{NON-SCI}}$ over time in Figure 11B for the Leave_Science group reflects learning or simply what students would have anticipated conditional on changing their focus from Science to other areas. However, in contrast, if the causal effect on interest of taking courses in a major is positive, a decrease in interest for Major j over time can reasonably be viewed as evidence that a student has learned that he is not as interested as expected in j. Thus, the decline in $\text{INTEREST}_{t,i,\text{SCI}}$ over time in Figure 11A for the Leave_Science group suggests that this group has learned that they were not as interested as expected in Science. Then, in terms of interest, it seems reasonable to conclude that students are, to some extent, pushed out of Science, but that it is difficult to determine whether or not they are also being pulled towards other majors. As discussed earlier, changes in interest in Science may be caused by changes in beliefs about academic performance in Science.

Heterogeneity within the Start_Other group

It would also be worthwhile to examine heterogeneity within each of the Stay_Science, Leave_Science, and Start_Other groups. Given the relatively small number of students in the Leave_Science and Stay_Science groups, we focus on the Start_Other group. Examining heterogeneity in beliefs within this group is useful, for example, for understanding why we found in Section III.1 (Figure 4A) that very few students in this group change into the Science major.

Students in the Start_Other group can be exposed to a small amount of science as part of the general studies curriculum, but are likely to predominantly take electives outside of the science area. Then, of particular interest is whether a student can learn that he/she may be talented at science under these circumstances. To this end, we examine the size of updates to $E(\text{AGPA}^t_{i,\text{SCI}})$ starting with changes that take place between the beginning of the first year (t=1) and the middle of the first year (t=2). Column 1 of the top panel of Table 1 shows that the sample average $E(\text{AGPA}^2_{i,\text{SCI}})-E(\text{AGPA}^1_{i,\text{SCI}})$ is −.17 for the 486 individuals in the Start_Other group who reported legitimate values at both t=1 and t=2 so that students tend to revise belief about performance/ability in Science downwards when they are not focusing on Science. However, the standard deviation of the update in the sample is relatively large, .68, and Column 1 of Table 1 shows that 28% of students in the Start_Other group have positive updates.

To understand why these positive updates do not lead to more changes into Science, we disaggregate further in Columns 2 and 3 of the first panel of Table 1 by stratifying on whether a person was in the bottom quartile (Column 2) or top three quartiles (Column 3) in terms of his beliefs at entrance, $E(AGPA^1_{i,SCI})$. Columns 2 and 3 reveal that the positive updating tends to be concentrated to a large extent in the (former) group of students who had very low initial expectations. For example, the sample average value of $E(AGPA^2_{i,SCI})$-$E(AGPA^1_{i,SCI})$ is .30 for students in the bottom quartile and -.34 for students in the top three quartiles. Over half of students in the bottom quartile had positive updates while only .19 of students in the top quartile had positive updates. Thus, the positive updating tends to take place primarily within a group which is not close to the margin of choosing Science; even after the updating, the student in the bottom quartile have a sample average value of $E(AGPA^2_{i,SCI})$, 2.24, which is almost a full point lower than its average value of $E(AGPA^2_{i,NON\text{-}SCI})$, 3.26. For the top 75% of students who likely begin much closer to the margin (e.g., sample average $E(AGPA^1_{i,SCI})$=3.34 and sample average $E(AGPA^2_{i,NON\text{-}SCI})$ =3.48), updating leads them to be further from the margin on average (e.g., sample average $E(AGPA^2_{i,SCI})$=2.91 and sample average $E(AGPA^2_{i,NON\text{-}SCI})$ =3.39). Columns 3-6 of the first panel of Table 1 show similar results when we examine updating between t=1 and t=6. For example, .66 of students in the bottom quartile in terms of $E(AGPA^1_{i,SCI})$ have positive updates between t=1 and t=6, but only .17 of students in the top three quartiles have positive updates. The second panel of Table 1 shows results that are similar to those in the first panel when we examine what a person has learned about $[E(AGPA^2_{i,SCI})-E(AGPA^1_{i,SCI})]-E(AGPA^2_{i,NON\text{-}SCI})-E(AGPA^1_{i,NON\text{-}SCI})$, his grade performance in science relative to his grade performance in other disciplines.

Thus, the evidence suggests that, while some people do update positively, learning that a person is very skilled is rare when students are not focusing specifically on Science. As discussed in the Conclusion, this could be either because the skills in Science are rather specific, in which case little may be learned about these skills while taking general classes, or because students who start in areas other than Science tend not to be overly skilled in Science.

V. Quantifying the importance of learning about performance/ability and other factors

The descriptive evidence in Section IV suggests that learning about academic performance/ability in Science is likely to play a very important role in determining whether a student chooses Science as his final major. In this Section we estimate simple models of college major choice. Because our models are reduced form in nature, we leave aside certain fundamental questions about, for example, the strategy students take after entrance in an effort to find a major with a good match. Instead our objective is much more modest - to provide a rough sense of the quantitative importance that learning about academic performance/ability and other factors play in the decision to major in Science.

As described in the previous section, our data contain information about the dependent variable of interest, college major, at each semester. Here we estimate the parameters of equation (1) by taking advantage of the data in the first (t=1) and last (t=6) semesters in our data. Unless otherwise stated, we focus on the 323 students who are still in school at t=6 and have no missing information at either t=1 or t=6.

Estimates from t=6

We begin with t=6. To a rough approximation, at this time students are certain about their final major so that we think of t=6 as corresponding to t* from Section II and use the stated major at t=6 as the dependent variable. As such, the analysis follows a standard discrete choice framework which is described by Equation (1)-(3) under the assumption that $\epsilon_{i,j}$ in equation (1) has an Extreme

Value distribution. We specify j=ED as the base case, X_i={$MALE_i$, $Math\_ACT_i$,$Verbal\_ACT_i$}, and $M_{i,j}$={$AGPA_{i,j}$, $AINCOME_{i,j}$} so that $E(M^{t*}_{i,j})$={$E(AGPA^6_{i,j})$, $E(AINCOME^6_{i,j})$}. Estimates of α_j and β are shown in Table 2.

As expected, the results show that, from both a statistical and quantitative standpoint, $AGPA_{i,j}$ is an especially important determinant of whether a person chooses major j. With respect to the former, the estimated effect and standard error of 3.22 and .27, respectively, imply a t-statistic in excess of close to twelve. With respect to the latter, the point estimate implies that a .50 increase in $AGPA_{i,j}$ in major j changes the odds ratio (the ratio of the probability j to the probability of an alternative k) by a factor of $e^{3.22x.50}$=5.0. Computing predicted probabilities using the first column of Table 2 can provide a sense of the prominent role of learning about grade performance/ability. The average predicted probability of choosing Science increases by 68% (from .117 to .197) under the counterfactual assumption $E(AGPA^6_{i,j})=E(AGPA^1_{i,j})$ for all j (i.e., that the mean of the distribution describing beliefs about GPA did not change during school for each j).[24] Given the reduced form nature of the empirical work, we do not think it is prudent to focus too much attention on the exact number from this calculation, but it does seem natural to conclude that the choice of Science is influenced very strongly by what a person learns about his performance/ability in Science.

Although the results point to $AGPA_{i,j}$ playing a particularly prominent role, consistent with the results of Arcidiacono et. al (2010) we also find evidence that, from both a statistical and quantitative standpoint, $AINCOME_{i,j}$ is also an important determinant of major choice. With respect to the former, the estimated effect and standard error of .056 and .008, respectively, imply a t-statistic in excess of six. With respect to the latter, the point estimate implies that a $5,000 increase in $AINCOME_{i,j}$ in major j changes the odds ratio (the ratio of the probability j to the probability of an alternative k) by a factor of $e^{.056x5}$=1.32. The average predicted probability of choosing science increases by 21% (from .117 to .142)

[24]For this sample, the average perceived probability is .118 at t=1 and .182 at t=6.

under the counterfactual assumption that $E(AINCOME^{6}_{i,j}) = E(AINCOME^{1}_{i,j})$ for all j.

Given our particular interest in Science, a desirable simplification for much of the analysis that follows involves, as in Figures 7A-11B, collapsing the set of non-science alternatives into a single NON-SCIENCE major so that choice set becomes {Science, Non-Science}. We find that this binary specification yields results that are similar in spirit to those in the uncollapsed specification. The first column of Table 2 shows that the coefficients associated with $AGPA_{i,j}$ and $AINCOME_{i,j}$ remain similar and size (2.65 vs. 3.22 and .048 vs. .056) and both remain statistically significant (t-statistics of 4.85 and 3.31 respectively). Our results quantifying the importance of learning also produce similar results. We find that the average predicted probability of choosing science increases by 58% (from .117 to .185) under the counterfactual assumption $E(AGPA^{6}_{i,j})=E(AGPA^{1}_{i,j})$ and for all j and the average predicted probability of choosing science increases by 23% (from .117 to .144) under the counterfactual assumption that $E(AINCOME^{6}_{i,j}) = E(AINCOME^{1}_{i,j})$ for all j.

Estimates from t=1

We could also perform a similar exercise by estimating the coefficients using information about the dependent variable from t=1. At t=1... (to be completed).

Thus, a person's beliefs about grade performance is a strong predictor of major. In order to attribute the estimated effect of $AGPA_{i,j}$ to $AGPA_{i,j}$ per se, it must be the case that there are no unobserved variables of importance that are both correlated with $AGPA_{i,j}$ and not caused by $AGPA_{i,j}$. One variable that might raise concern is a person's INTEREST. Earlier we found no evidence that a person's interest in Science had a causal effect on the person's beliefs about academic performance in Science, at least to the extent that a person's effort is the logical avenue through which such an effect must arise.[25] Then, the logical remaining concern is that a person's interest in Science and a person's beliefs about GPA in Science are both influenced by some third common factor. Seemingly the most logical common factor of this type is a background factor such as whether a person's parents work in a science occupation. If such factors are important, one would expect students with higher levels of interest in Science at t=1 to continue to have higher levels of interest in Science at t=6 (than students with lower levels of interest at t=1) even if they learn that their grade performance/ability in Science is lower than expected. Then, Figure 11A does not provide much evidence in support of this concern since $INTEREST_{t,i,SCI}$ for the Leave_Science group looks very similar to $INTEREST_{t,i,SCI}$ for the Start_Other group after the former group realizes that their grade performance will not be particularly high. Thus, it seems that the relationship between a person's interest in Science and a person's beliefs about GPA in Science is, at least to a large extent, generated by GPA having a causal effect on Interest.

Further, even if a common factor causes a correlation between, for example, $INTEREST_{t,i,SCI}$ and $E(AINCOME^{t}_{i,SCI})$, it is not evident that there should also exist a correlation between changes in $INTEREST_{t,i,SCI}$ and changes in $E(AINCOME^{t}_{i,SCI})$. This would suggest the benefit of examining the robustness of our results to models in which changes in beliefs about M_{ij} enter directly. As discussed in S&S (2009), from a conceptual standpoint, such models are based on the notion that Science will be chosen if what the person learns about the expected benefits of Science relative to the other majors is sufficient to push the student into Science (or push the student out of Science) given how close to the margin of indifference he was at the time of entrance. Conceptually, this model is most straightforward when the dependent variable is whether or not to choose Science. In this binary model, our specification includes both variables that measure learning directly (i.e., [$E(AGPA^{6}_{i,SCI})$-

[25] One could estimate the model using $ABILITY_{i,j}$ rather than $AGPA_{i,j}$. The estimated effect remains significantly statistic despite the fact that the $ABILITY_{i,j}$ measure is likely to be quite noisy, (t-statistic=5.80).

$E(AGPA^6_{i,NON\text{-}SCI})$]-[$E(AGPA^1_{i,SCI})$-$E(AGPA^1_{i,NON\text{-}SCI})$] and [$E(INCOME^6_{i,SCI})$-$E(INCOME^6_{i,NON\text{-}SCI})$]-[$E(AINCOME^1_{i,SCI})$-$E(AINCOME^1_{i,NON\text{-}SCI})$]) and also variables that help characterize how positive a person was about Science at the time of entrance (i.e., $E(AGPA^1_{i,SCI})$-$E(AGPA^1_{i,NON\text{-}SCI})$, $E(AINCOME^1_{i,SCI})$-$E(AINCOME^1_{i,NON\text{-}SCI})$, and Pr^1(Science).

In this specification we find an estimated effect (std. error) of 3.627 (1.00) for [$E(AGPA^6_{i,SCI})$-$E(AGPA^6_{i,NON\text{-}SCI})$]-[$E(AGPA^1_{i,SCI})$-$E(AGPA^1_{i,NON\text{-}SCI})$]. Consistent with our earlier findings, we find that the average predicted probability of choosing Science increases by 54% (from .117 to .181) under the counterfactual assumption that no learning takes place about academic performance (i.e., setting [$E(AGPA^6_{i,SCI})$-$E(AGPA^6_{i,NON\text{-}SCI})$]-[$E(AGPA^1_{i,SCI})$-$E(AGPA^1_{i,NON\text{-}SCI})$]=0 for all students.

VI. Conclusion

References

Altonji, Joseph. “The Demand for and Return to Education When Education Outcomes are Uncertain,” Journal of Labor Economics, 1993, vol. 11, no. 1, 48-83.

Arcidiacono, Peter. “Ability Sorting and the Returns to College Major,” Journal of Econometrics, Vol. 121, Nos. 1-2 (August, 2004), 343-375.

Arcidiacono, Peter, Hotz, Joseph, and Kang, Songman, “Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals,” Duke University working paper, 2010.

Blass, Asher, Lach, Saul, and Manski, Charles. “Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability.

COSEPUP (Committee on Science, Engineering, and Public Policy), “Rising Above the Gathering Storm: Engergizing and Employing America for a Brighter Economic Future. The National Academies Press (2007).

Dominitz, Jeff. “Earnings Expectations, Revisions, and Realizations,” The Review of Economics and Statistics, August 1998, 374-388.

Dominitz, Jeff and Manski, Charles. “Eliciting Student Expectations of the Returns to Schooling,” Winter 1996, 1-26.

Dominitz, Jeff and Manski, Charles. “Using Expectations Data to Study Subjective Income Expectations,” Journal of American Statistical Association, September 1997, 855-867.

Dominitz, Jeff and Manski, Charles. “How Should We Measure Consumer Confidence (sentiment)? Evidence form the Michigan Survey of Consumers,” Working paper 9926, National Bureau of Research, 2003.

Juster, T., “Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design,” Journal of hte American Statistical Association 61 (1966), 658-96.

Manski, C., “The Use of Intentions Data to Predict Behavior: A Best Case Analysis,” Journal of the American Statistical Association 85 (1990), 934-940.

Manski, C., “Measuring Expectations,” Econometrica, 72 (5), (2004), 1329-1376.

Montmarquette, C, Cannings, Kathy, and Mahseredjian, Sophie, “How Do Young People Choose College Majors?,” Economics of Educaiton Review, Elsevier, 21(6), (2002) 543-556, December.

Stinebrickner, Todd and Stinebrickner, Ralph, “Time-Use and College Outcomes,” Journal of Econometrics, 121 (1-2) July-August (2004), 243-269.

Stinebrickner, Todd and Stinebrickner, Ralph, “The Effect of Credit Constrains on the College Drop-Out Decision: A Direct Approach Using a New Panel Study,” American Economic Review. December (2008a)

Stinebrickner, Todd and Stinebrickner, Ralph, “The Causal Effect of Studying on Academic Performance,” Frontiers in Economic Policy and Analysis (*Frontiers*), Berkeley Electronic Press (2008b).

Stinebrickner, Todd and Stinebrickner, Ralph, “Learning about Academic Ability and the College Drop-Out Decision (2009).

Table 1

	Changes in EGPA - Start_Other group					
	t=2 full sample n=486	t=2 bottom 25% n=131	t=2 top 75% n=355	t=6 full sample n=334	t=6 bottom 25% n=83	t=6 top 75% n=251
$E(AGPA^{t}_{i,SCI})-E(AGPA^{1}_{i,SCI})$	-.17 (.68)	.30 (.71)	-.34 (.58)	-0.21	.45 (.76)	-.43 (.58)
Proportion with $E(AGPA^{t}_{i,SCI})-E(AGPA^{1}_{i,SCI})>0$	0.28	0.51	0.19	0.29	0.66	0.17
$E(AGPA^{1}_{i,SCI})$	2.97 (.76)	1.93 (.56)	3.35 (.36)	2.99 (.75)	1.91 (.59)	3.34 (.37)
$E(AGPA^{t}_{i,SCI})$	2.80 (.75)	2.24 (.68)	3.01 (.66)	2.77 (.61)	2.37 (.56)	2.91 (.56)
$E(AGPA^{1}_{i,NON\_SCI})$	3.41 (.338)	3.25 (.35)	3.48 (.31)	3.42 (.32)	3.26 (.34)	3.48 (.30)
$E(AGPA^{t}_{i,NON\_SCI})$	3.36 (.347)	3.26 (.35)	3.40 (.33)	3.36 (.34)	3.27 (.35)	3.39 (.32)
	t=2 full sample n=486	t=2 bottom 25% n=121	t=2 top 25% n=365	t=6 full sample n=334	t=6 bottom 25% n=82	t=6 top 25% n=252
$[E(AGPA^{t}_{i,SCI})-E(AGPA^{t}_{i,NON-SCI})]$	-.11 (.69)	.34 (.80)	-.27 (.57)	-.145 (.69)	.47 (.71)	-.34 (.55)
Proportion with $[E(AGPA^{t}_{i,SCI})-E(AGPA^{t}_{i,NON-SCI})]-[E(AGPA^{1}_{i,SCI})-E(AGPA^{1}_{i,NONSCI})]>0$	0.38	0.66	0.29	0.39	0.75	0.27
$E(AGPA^{1}_{i,SCI})-E(AGPA^{1}_{i,NON-SCI})$	-.44 (.70)	-1.43 (.55)	-.11 (.57)	-.43 (.70)	-1.41 (.58)	-.11 (.353)
$E(AGPA^{t}_{i,SCI})-E(AGPA^{t}_{i,NON-SCI})$	-.56 (.69)	-1.08 (.71)	-.39 (.59)	-.58 (.55)	-.93 (.50)	-.46 (.52)

Notes: The descriptive statistics in the Table show mean (standard deviation).
In first panel:Bottom 25% refers to students with $E(AGPA^{1}_{i,SCI})<2.5$ and Top 75% refers to students with $E(AGPA^{1}_{i,SCI})>2.5$. In second panel: Bottom 25% refers to students with $E(AGPA^{1}_{i,SCI})-E(AGPA^{1}_{i,NON-SCI})<-.795$ and Top 75% refers to students with $E(AGPA^{1}_{i,SCI})-E(AGPA^{1}_{i,NON-SCI})>-.795$.

Table 2 Major Choice

	Logit	
	t=6 Full choice set	t=6 Binary Choice Set
$AGPA_{i,j}$	3.223 (.274)*	2.659 (.548)*
$AINCOME_{i,j}$	.056 (.008)*	.048 (.014)*
SCIENCE		
Constant	1.309 (1.820)	-2.358 (1.273)
Male	.241 (.589)	-.022 (.419)
Math_ACT	.080 (.077)	.098 (.055)
Verbal_ACT	-.099 (.074)	-.049 (.053)
AG		
Constant	2.203 (2.00)	
Male	-.197 (.648)	
Math_ACT	-.122 (.086)	
Verbal_ACT	.025 (.080)	
BUS		
Constant	3.398 (1.809)	
Male	.346 (.571)	
Math_ACT	.070 (.079)	
Verbal_ACT	-.218 (.075)*	
ED		
Base Case	N.A.	
HUM		
Constant	.634 (1.66)	
Male	.709 (.524)	
Math_ACT	-.044 (.067)	
Verbal_ACT	.053 (.067)	
PRO		
Constant	1.309 (1.820)	
Male	.241 (.589)	
Math_ACT	.080 (.077)	
Verbal_ACT	-.099 (.074)	
SS		
Constant	3.139 (1.690)	
Male	-.307 (.559)	
Math_ACT	.008 (.070)	
Verbal_ACT	-.114 (.067)	

Notes: * Significant at 5%

Table 1A

	Proportion stratified by observable characteristics						
	AG	BUS	ED	HUM	SCI	PROF	SS
	t=1						
Male n=129	0.116	0.151	0.045	0.195	0.281	0.08	0.132
Female n=242	0.1	0.126	0.09	0.192	0.16	0.185	0.165
Math_ACT <Median n=179	0.12	0.156	0.08	0.128	0.163	0.16	0.185
Math_ACT>Median n=192	0.06	0.115	0.069	0.253	0.238	0.135	0.124
Verbal_ACT<Median n=199	0.12	0.186	0.064	0.093	0.223	0.188	0.123
Verbal_ACT>Median n=172	0.06	0.075	0.086	0.309	0.177	0.101	0.189
	t=6						
Male	0.09	0.197	0.058	0.269	0.118	0.137	0.133
Female	0.1	0.09	0.103	0.219	0.109	0.25	0.14
Math_ACT<Median n=179	0.11	0.122	0.092	0.237	0.072	0.23	0.136
Math_ACT>Median n=192	0.06	0.132	0.083	0.235	0.149	0.196	0.139
Verbal_ACT<Median n=199	0.08	0.183	0.065	0.179	0.116	0.27	0.101
Verbal_ACT>Median n=172	0.09	0.063	0.113	0.302	0.107	0.145	0.18

Median Math ACT score is 22
Median Verbal ACT score is 24

Table 1B

	Average probability stratified by observable characteristics						
	AG	BUS	ED	HUM	SCI	PROF	SS
	t=1						
Male n=129	0.134	0.167	0.075	0.179	0.217	0.09	0.139
Female n=242	0.09	0.124	0.104	0.189	0.162	0.183	0.15
Math_ACT <Median n=179	0.133	0.159	0.1	0.154	0.153	0.159	0.151
Math_ACT>Median n=192	0.08	0.121	0.089	0.215	0.207	**0.141**	0.142
Verbal_ACT<Median n=199	0.128	0.187	0.09	0.107	0.194	0.173	0.12
Verbal_ACT>Median n=172	0.08	0.084	0.096	0.277	0.165	0.122	0.176
	t=6						
Male n=129	0.102	0.207	0.058	0.254	0.118	0.134	0.123
Female n=242	0.09	0.101	0.112	0.222	0.108	0.219	0.147
Math_ACT<Median n=179	0.111	0.135	0.1	0.237	0.075	0.204	0.135
Math_ACT>Median n=192	0.08	0.14	0.088	0.229	0.145	0.176	0.141
Verbal_ACT<Median n=199	0.09	0.189	0.799	0.181	0.116	0.232	0.104
Verbal_ACT>Median n=172	0.09	0.078	0.11	0.293	0.106	0.141	0.178

Note:As in all of Section III, t_1 is the beginning of the second semester so t_0:t_1 is the first semester.

Figure 1A Avg. perceived probability of semester t stated major

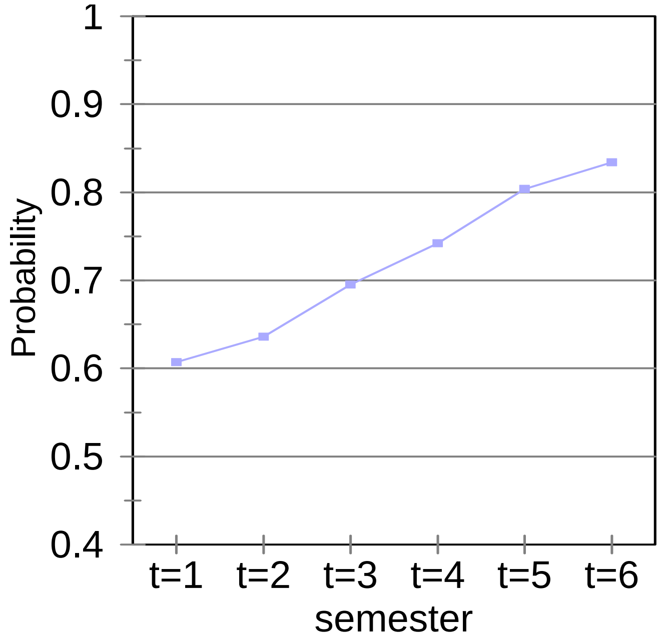


Figure 1B Avg. perceived probability of final major as of t

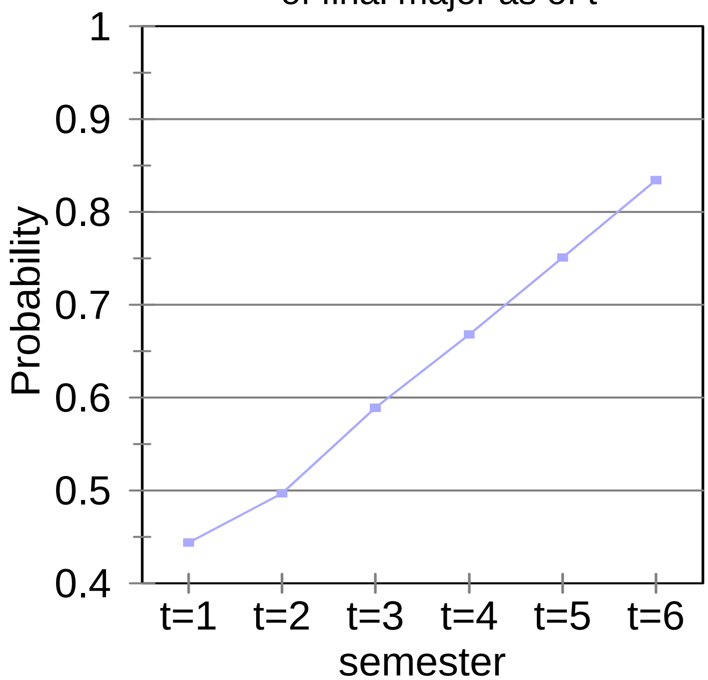


Figure 2A Proportion with stated major j in semester t

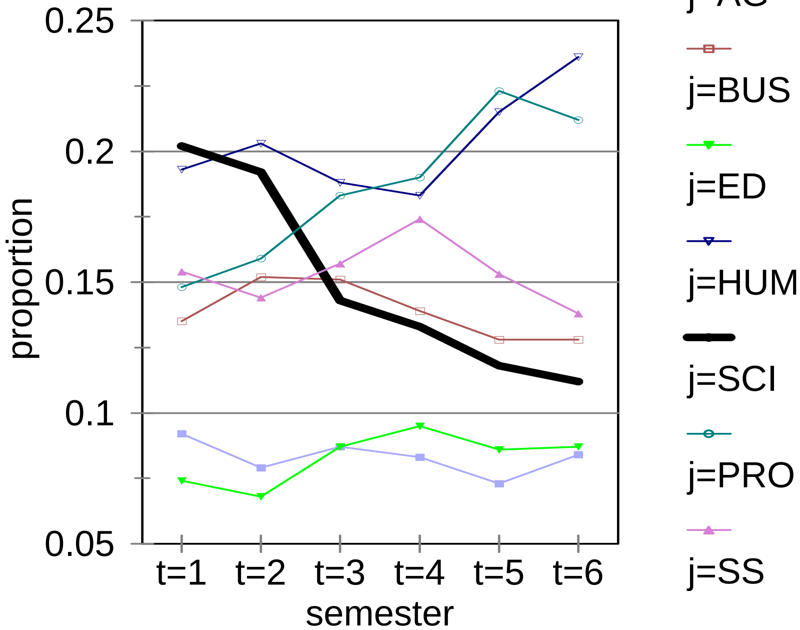


Figure 2B Avg. perceived probability of major j in semester t

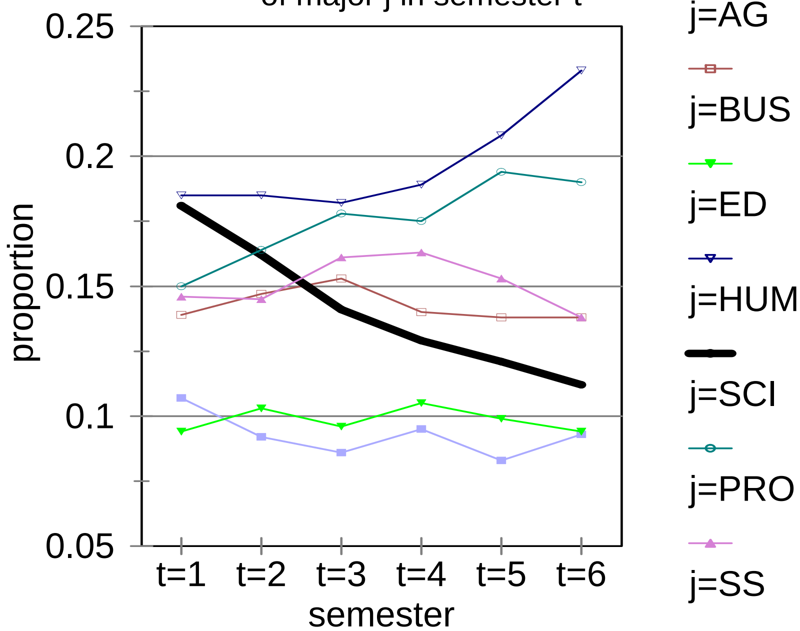


Figure 3A Actual probability of staying in starting major j

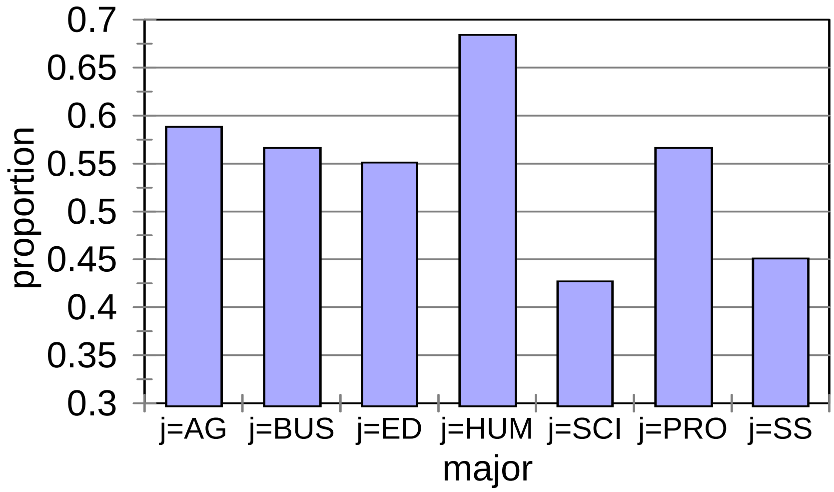


Figure 3B Avg. perceived probability of staying in starting major j

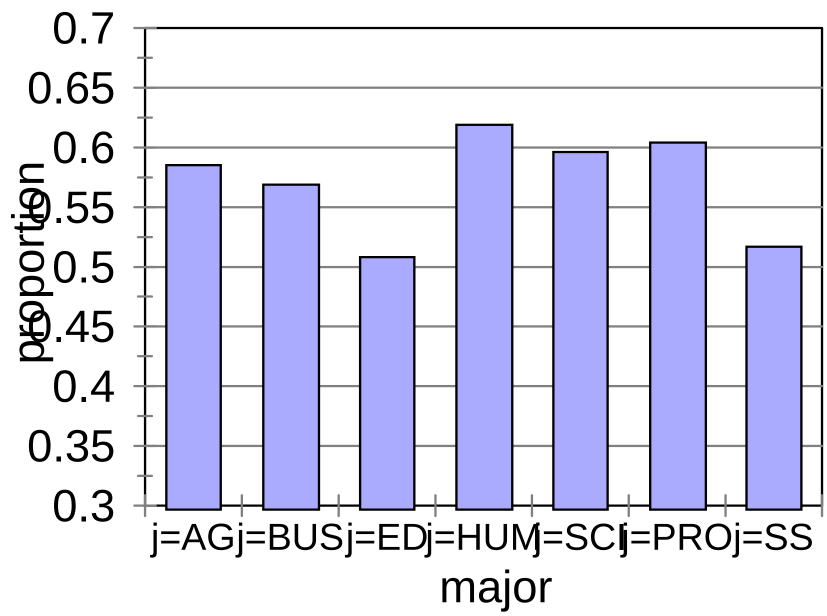


Figure 4A Actual probability
of changing to final major of j

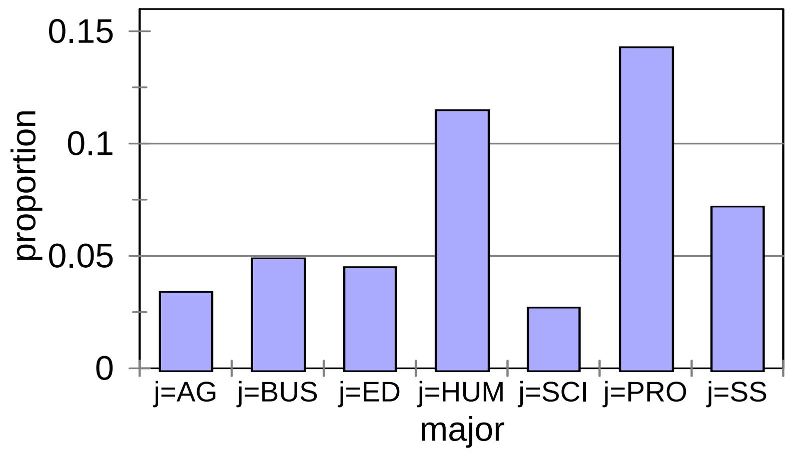
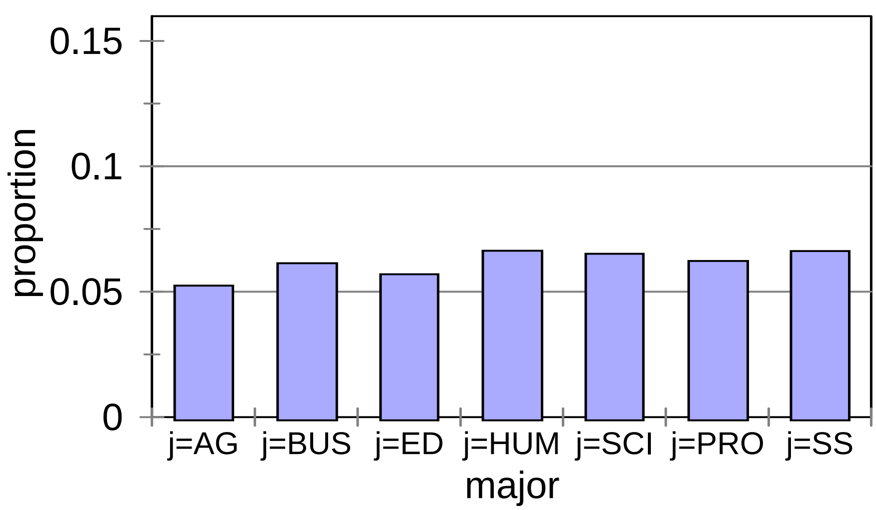


Figure 4B Avg. perceived probability
of changing to final major of j



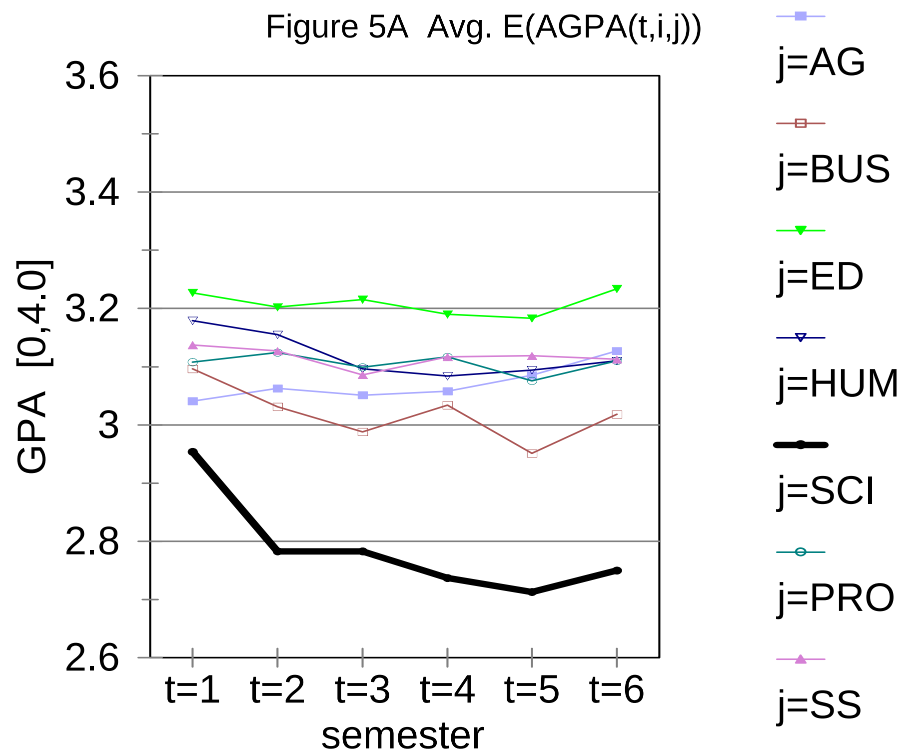


Figure 5A Avg. E(AGPA(t,i,j))
3.6
3.4
3.2
3
2.8
2.6
GPA [0,4.0]
t=1 t=2 t=3 t=4 t=5 t=6
semester
j=AG
j=BUS
j=ED
j=HUM
j=SCI
j=PRO
j=SS

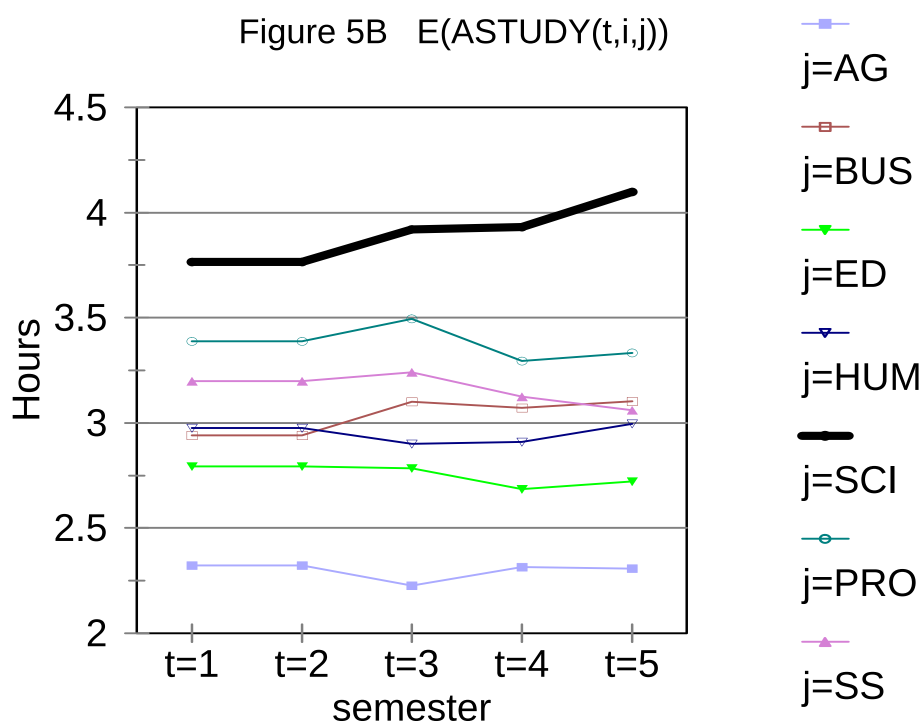


Figure 5B E(ASTUDY(t,i,j))
4.5
4
3.5
3
2.5
2
Hours
t=1 t=2 t=3 t=4 t=5
semester
j=AG
j=BUS
j=ED
j=HUM
j=SCI
j=PRO
j=SS

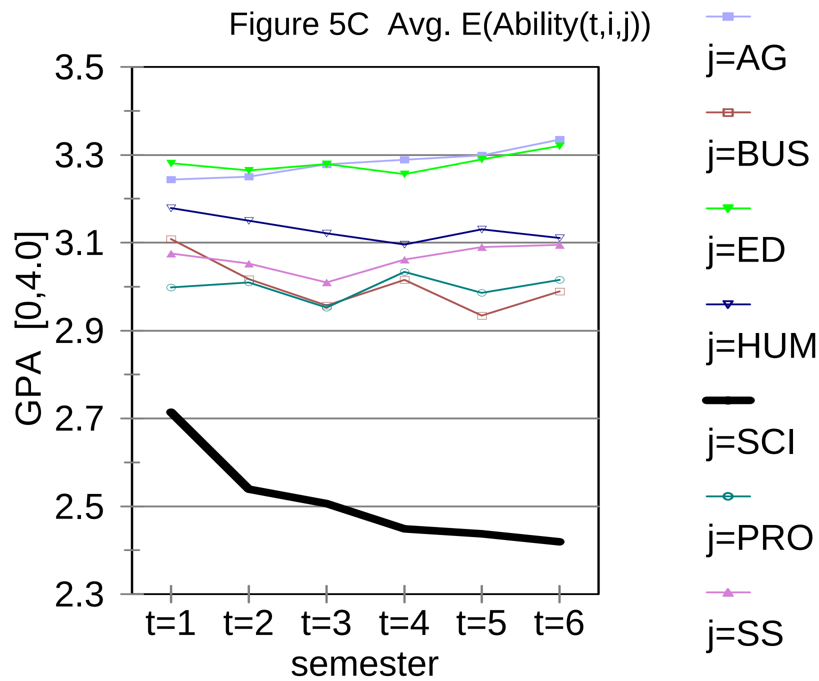


Figure 5C Avg. E(Ability(t,i,j))
GPA [0,4.0]
3.5
3.3
3.1
2.9
2.7
2.5
2.3
t=1
t=2
t=3
t=4
t=5
t=6
semester
j=AG
j=BUS
j=ED
j=HUM
j=SCI
j=PRO
j=SS

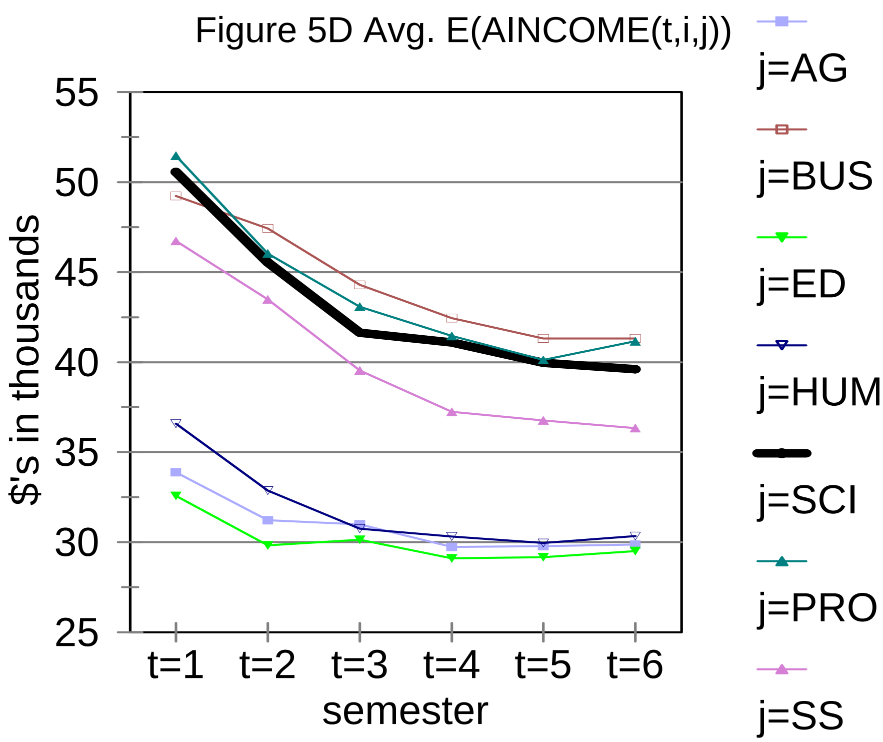


Figure 5D Avg. E(AINCOME(t,i,j))
$'s in thousands
55
50
45
40
35
30
25
t=1
t=2
t=3
t=4
t=5
t=6
semester
j=AG
j=BUS
j=ED
j=HUM
j=SCI
j=PRO
j=SS

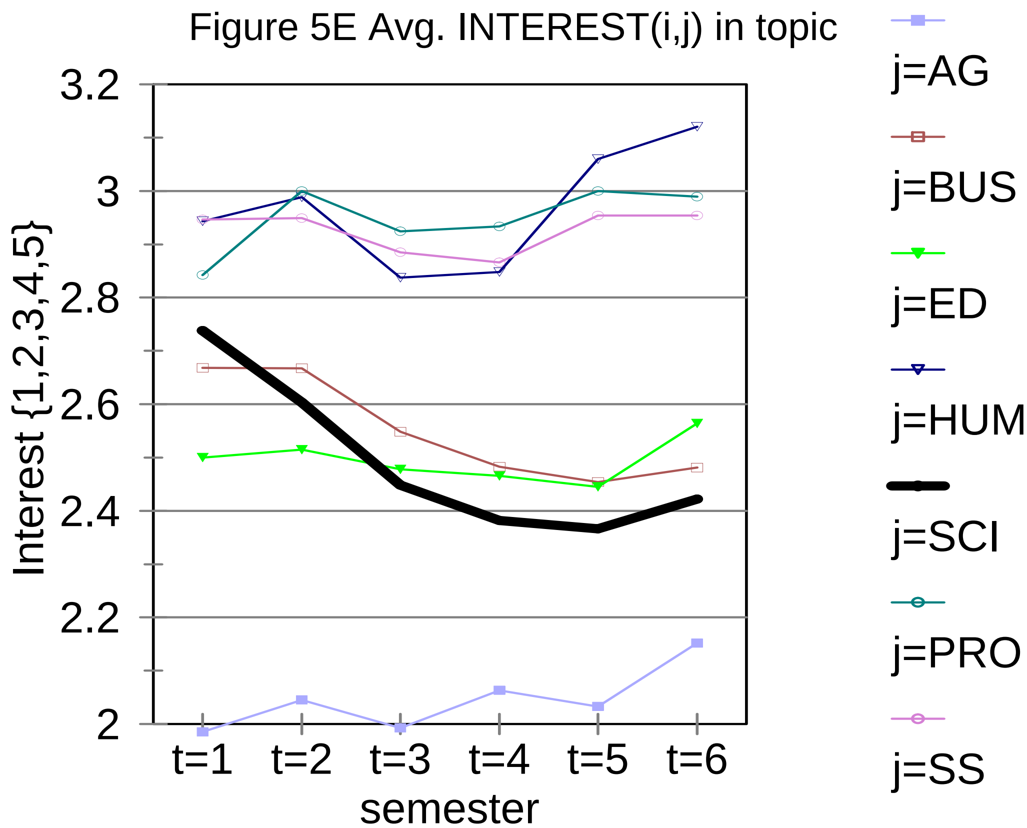


Figure 5E Avg. INTEREST(i,j) in topic
Interest {1,2,3,4,5}
3.2
3
2.8
2.6
2.4
2.2
2
t=1
t=2
t=3
t=4
t=5
t=6
semester
j=AG
j=BUS
j=ED
j=HUM
j=SCI
j=PRO
j=SS

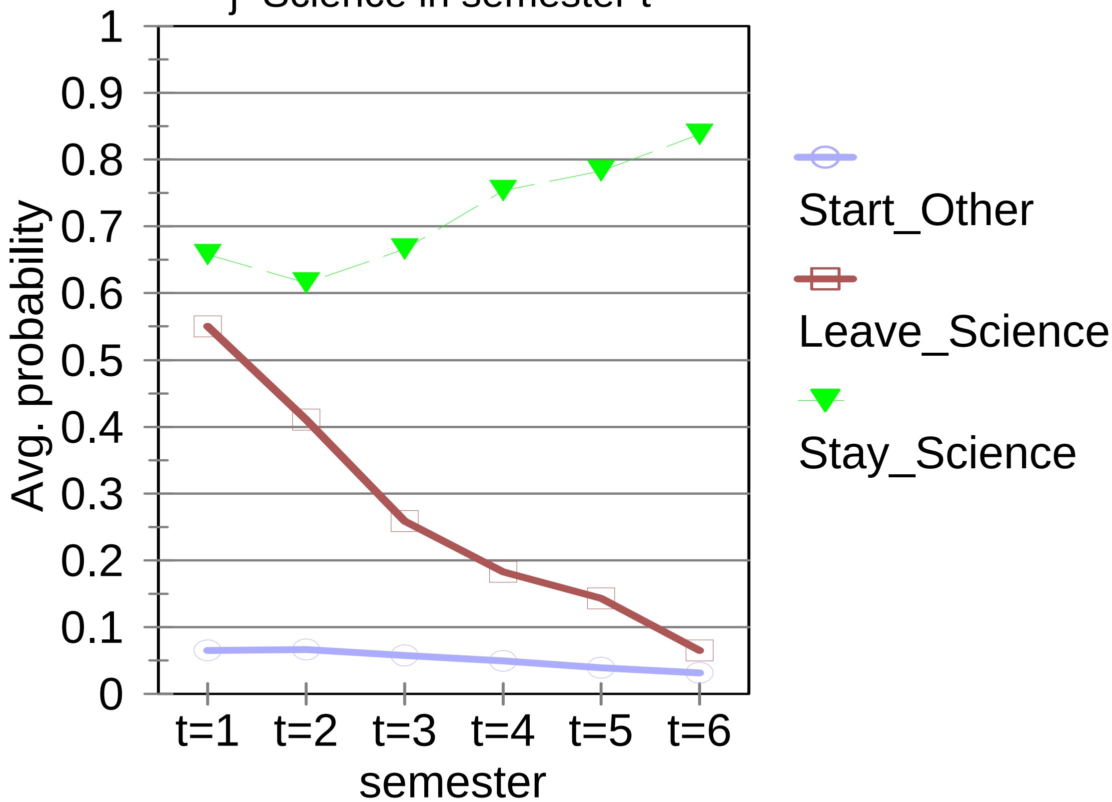


Figure 6 Avg. perceived probability of j=Science in semester t

Figure 7A E(AGPA(t,i,SCI))

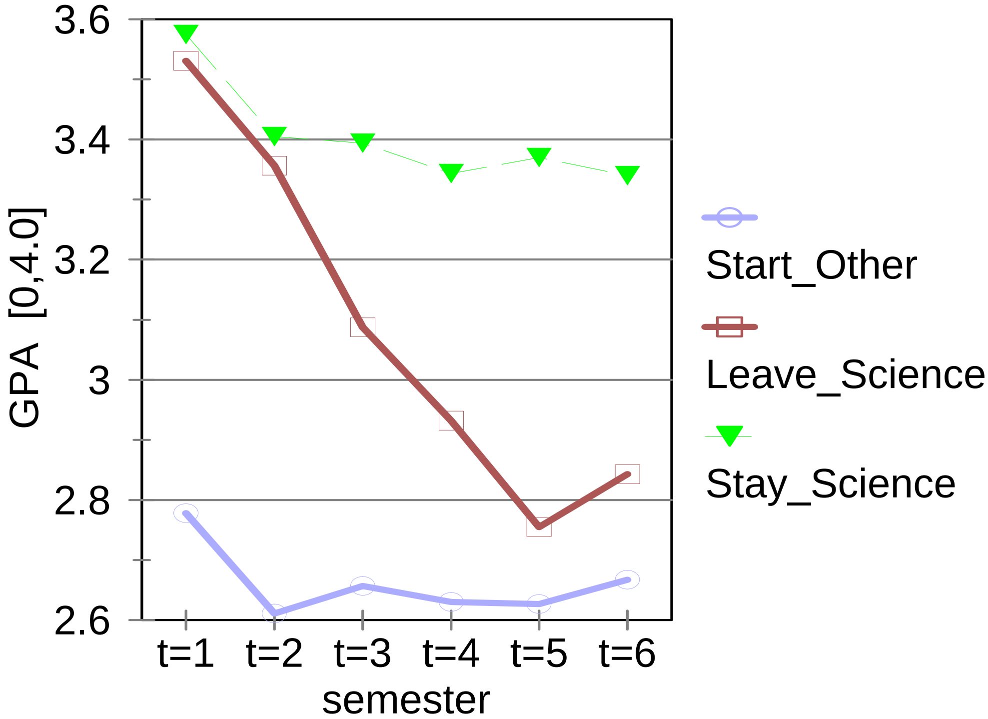


Figure 7B E(GPA(t,i,NON-SCI)

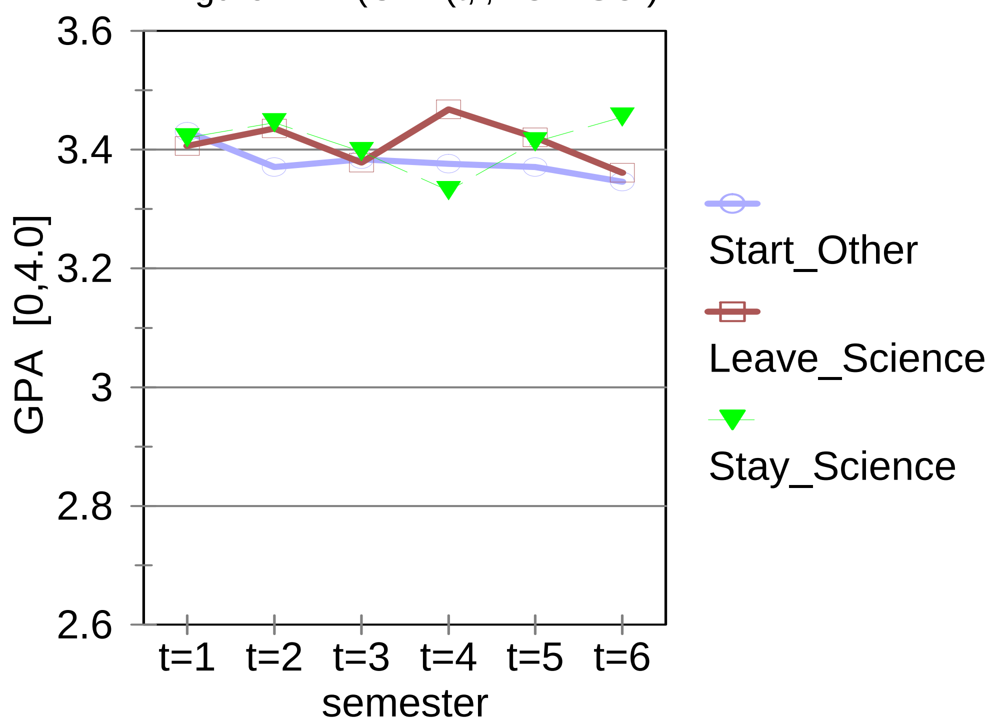


Figure 8A E(ASTUDY(t,iSCI))

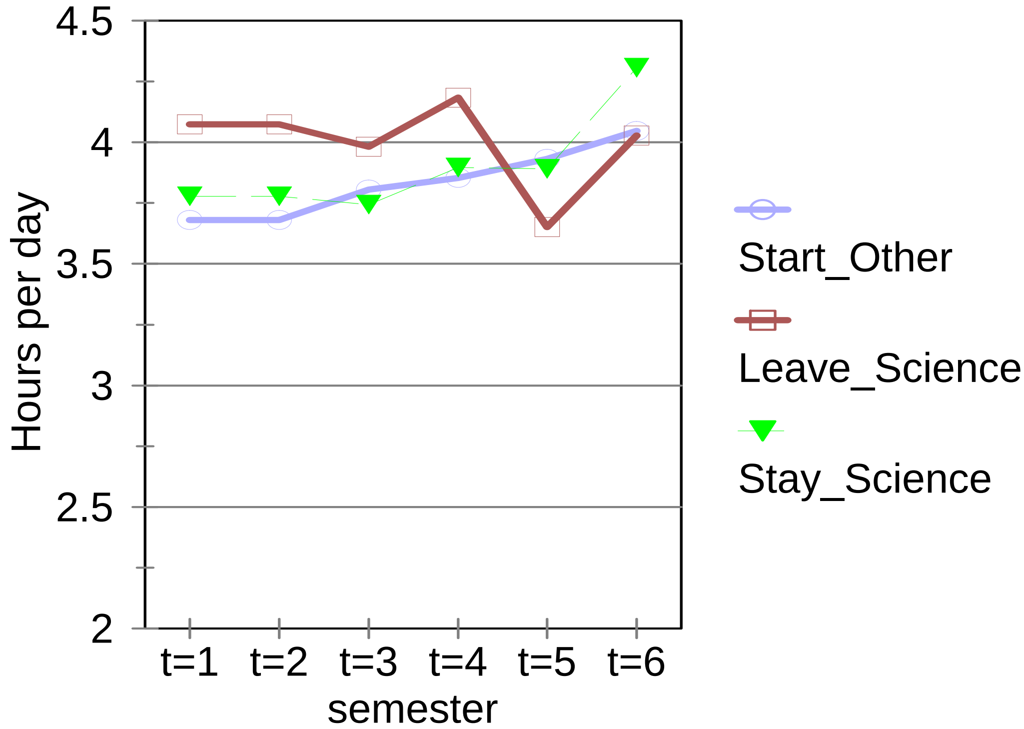


Figure 8B E(ASTUDY(t,i,NON-SCI))

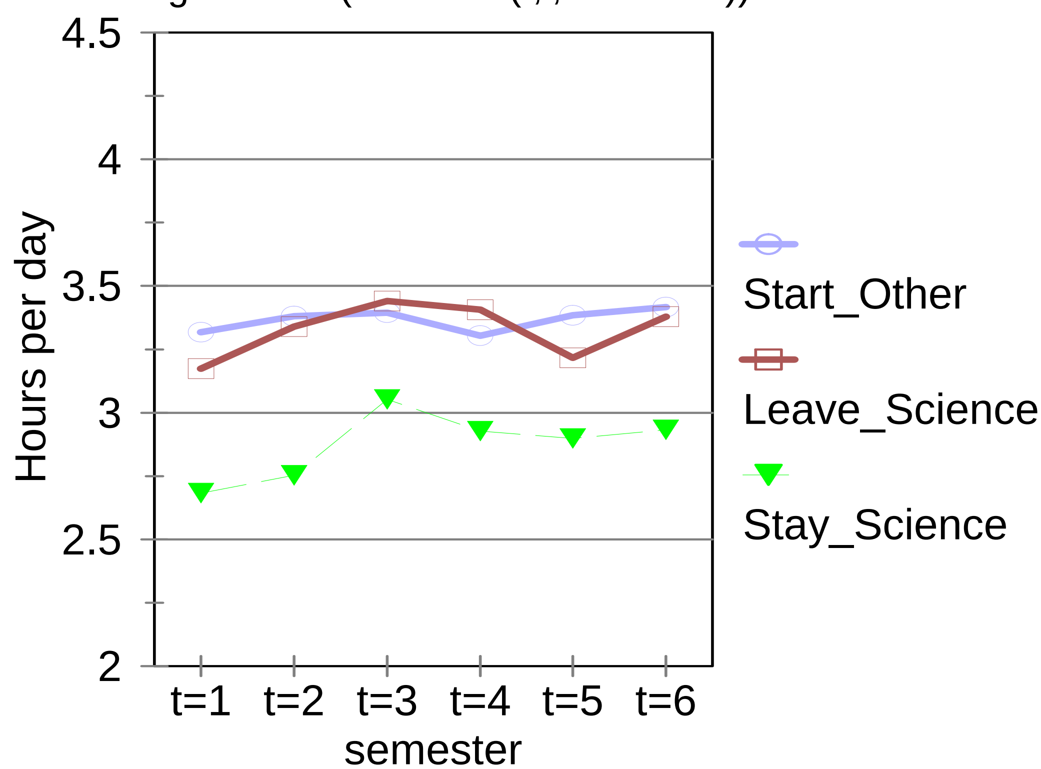


Figure 9A E(ABILITY(t,i,SCI))

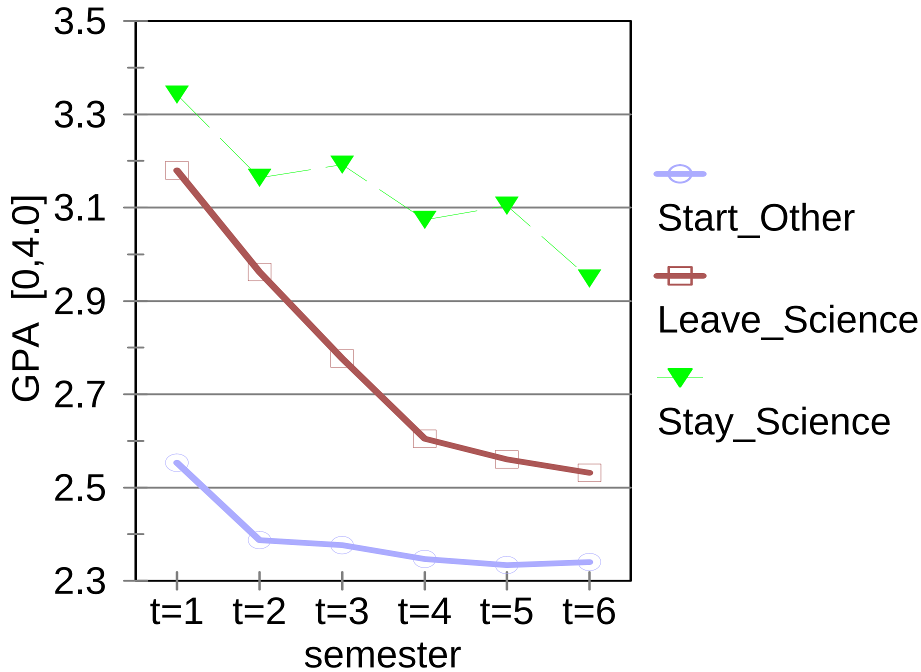


Figure 9B E(ABILITY(t,i,NON-SCI))

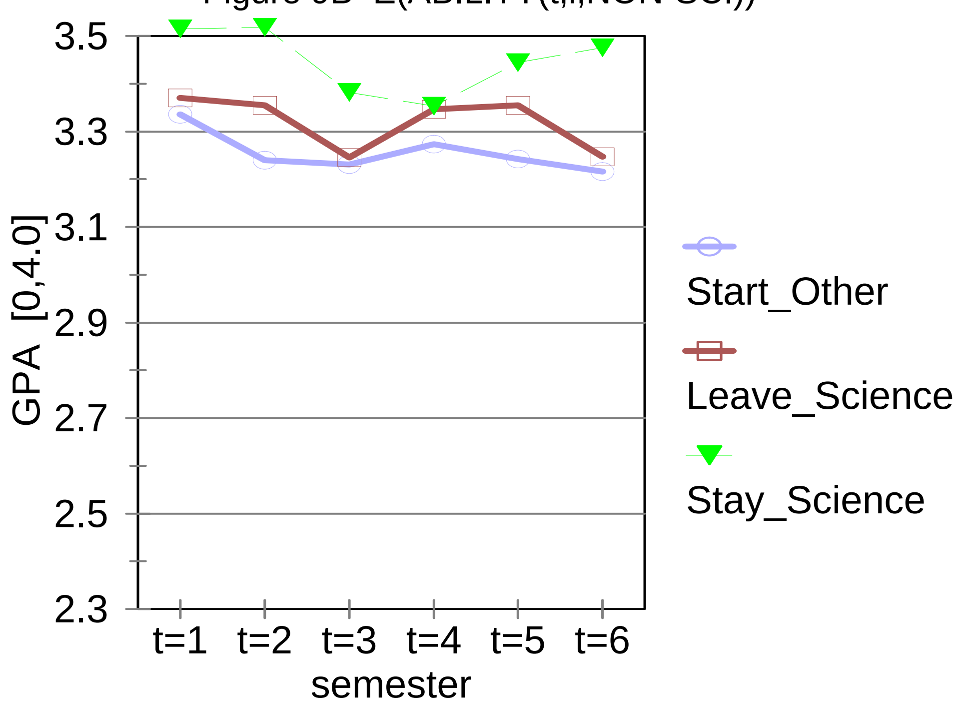


Figure 10A E(AINCOME(t,i,SCI)

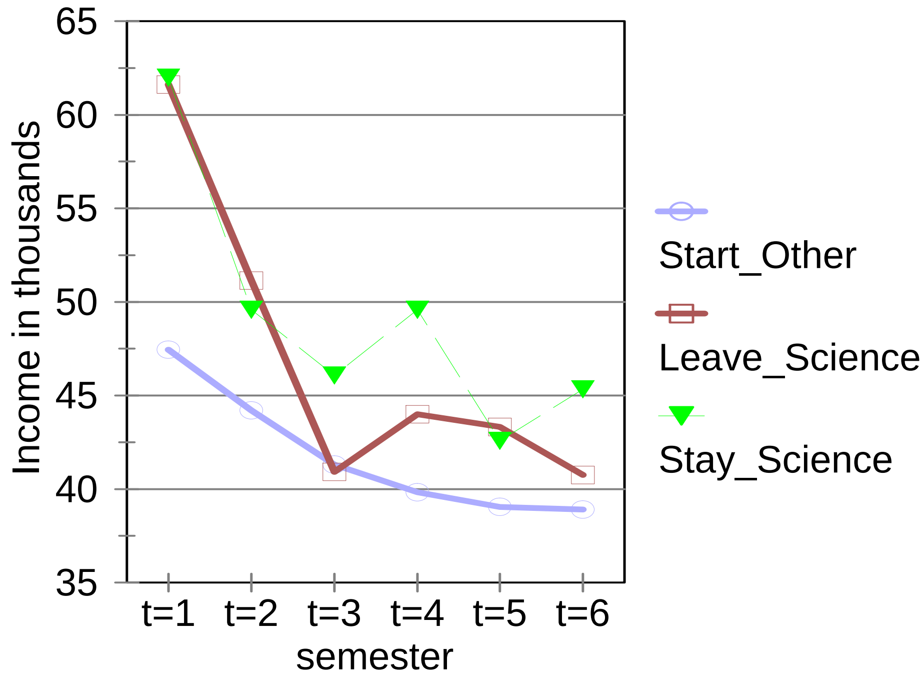


Figure 10B E(AINCOME(t,i,NON-SCI)

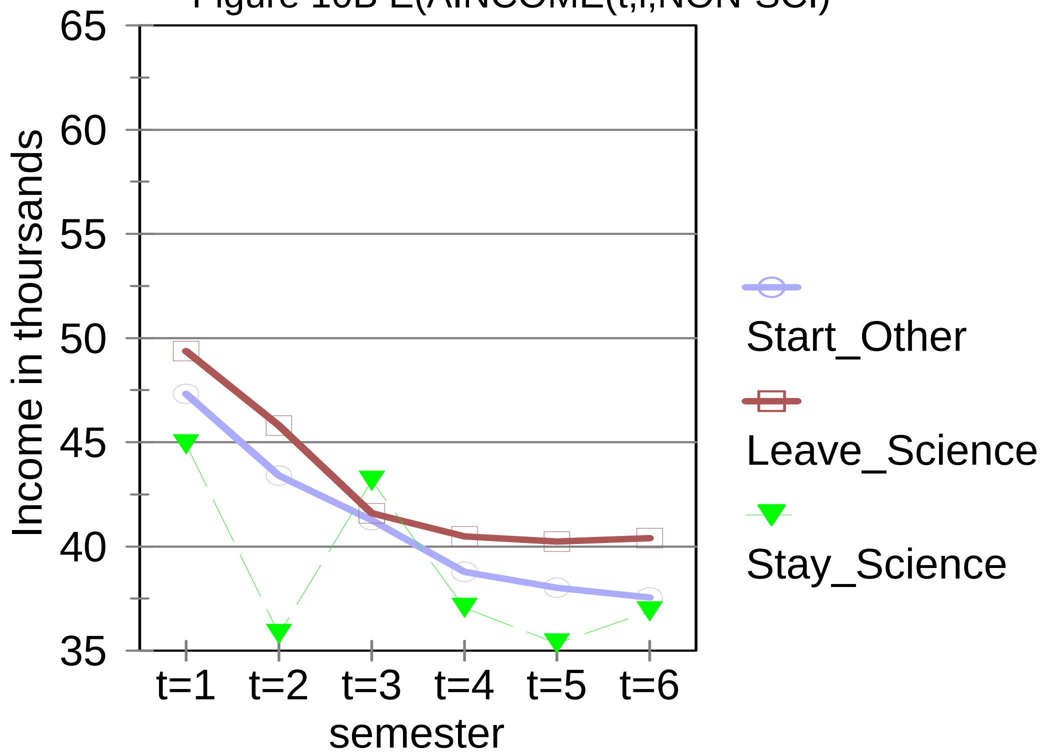


Figure 11A INTEREST(t,i,SCI)

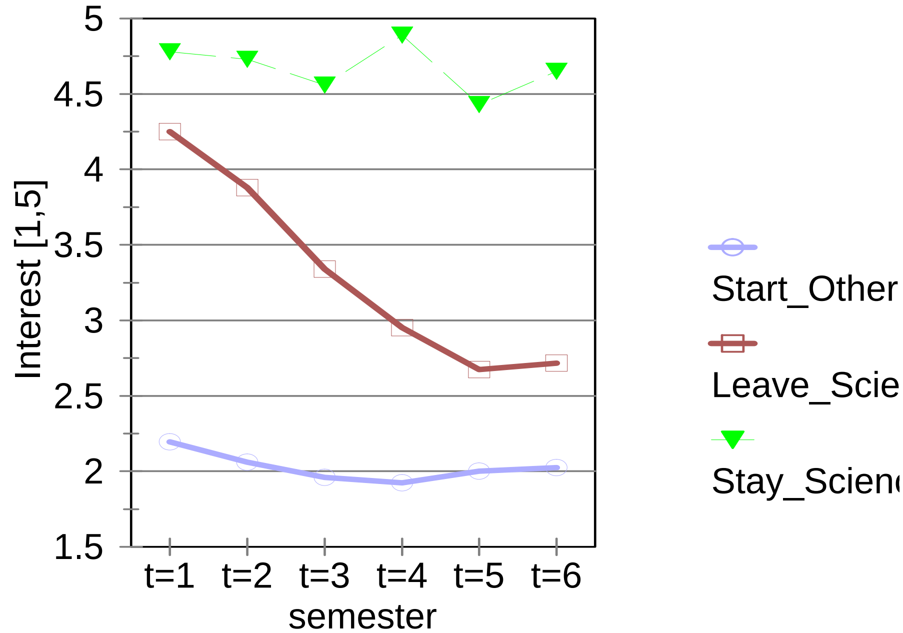
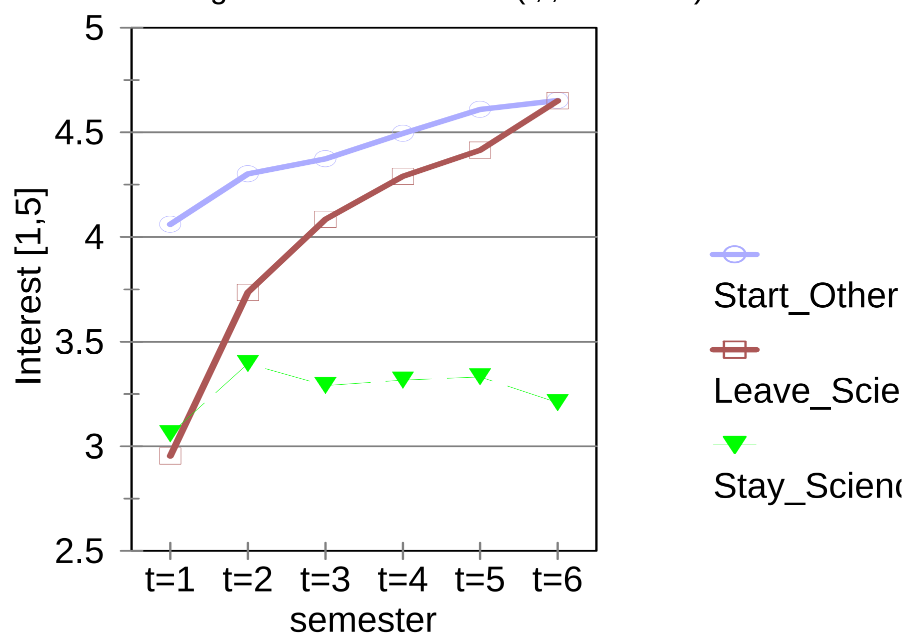


Figure 11B INTEREST(t,i,NON-SCI)



Appendix

Question 1. We realize that you may not be sure what area of study you will eventually graduate with. In the first column below are listed possible areas of study. In the second column write down the percent chance that you will end up with this area of study (note: the percent chance for each particular area of study should be between 0 and 100 and the numbers in the percent chance column should add up to 100). In the third column, please write down the grade point average (GPA) you would expect to receive in a typical semester in the future if you had each of these areas of study. In the fourth column write down the **yearly** income you would expect to earn at age 28 (or 10 years from now if you are now 20 years of age or older) if you graduated with each of these areas of study. In the fifth column, write down how interesting you find each particular area of study. In this column enter a number 1-5 where 1=extremely interested, 2=quite interested, 3=some interest, 4=very little interest, 5=not interested.

Please fill out all remaining columns even if you have a zero in the percent chance column for a particular area of study.

Humanities include Art, English, Foreign Languages, History, Music, Philosophy, Religion, and Theatre.

Natural Science and Math includes Biology, Chemistry, Computer Science, Physics and Mathematics.

Professional Programs include Industrial Arts, Industrial Technology, Child Development, Dietetics, Home Economics, Nutrition, and Nursing.

Social Sciences include Economics, Political Science, Psychology and Sociology.

*****When considering Expected GPA in an area of study consider ALL courses you will take if you have that area of study -including both courses that are required for your major and all other courses.*****

Area of study	Percent Chance (see above)	Expected GPA (0.00-4.00) **	Expected Yearly Income Age 28 (in dollars)	Interest Level in Area of Study Enter 1-5 1. Extremely interested 2. Quite interested 3. Some interest 4. Very little interest 5. Not interested
1.Agriculture (and Natural Resources)	_________	_________	__________	__________
2.Business	_________	_________	___________	___________
3.Elementary Education	_________	_________	___________	___________
4. Humanities	_________	_________	___________	___________
5.Natural Science & Math	_________	_________	___________	___________
6. Professional Programs	_________	_________	____________	___________
7. Social Sciences	__________	_________	____________	___________

Note: Numbers in the second column (percent chance) should each be between 0 and 100 and should add up to 100.
Note: A=4.0, B=3.0, C=2.0, D=1.0, F=0.0. So numbers in third column (GPA) should be between 0.00 and 4.00.

although the estimated effect is smaller (.879) and the The average predicted probability of choosing Science increases by 26% (from .120 to .152) under the counterfactual assumption $E(AGPA^{6}_{i,j})=E(AGPA^{1}_{i,j})$ for all j