Gender Differences in Persistence in a Field of Study

Michael Kaganovich*, Morgan Taylor†, Ruli Xiao‡

October 29, 2019

Preliminary version, not for circulation [1]

Extended Abstract

The underrepresentation of women in some quantitatively oriented academic fields such as STEM[2], business, and economics is widely seen across U.S. colleges and universities. Among potential explanations for this gender imbalance, some scholars have noted that the average grade levels in these fields tend to be substantially lower than in others, and have hypothesized that female students are generally characterized by relatively greater sensitivity to the grades they receive. This paper undertakes an examination of these questions using a rich Indiana University Learning Analytics dataset, which fully represents undergraduate student academic activities over the period of 2006 through 2017. After conditioning on a student's starting academic field, we estimate the gender effect on students' academic paths in terms of the likelihood of persistence in their original academic area or switching to a particular academic alternative. This analysis supports the conjectures that women in STEM, business, and economics fields do exhibit relatively stronger sensitivity to grades in their decisions about persisting or switching to another field, which in part helps explain their lower representation among graduates in these areas. Our significant new finding is that the conclusion about women's greater sensitivity to grades is specific to the aforementioned fields of study and does not universally extend to other starting academic disciplines. We find, in particular, that in the social sciences and humanities category, women demonstrate either equal or higher levels of persistence across various grade levels, relative to men. These empirical results suggest that stronger sensitivity to grades, rather than being a gender-specific phenomenon, is more likely to reflect gender differences in the underlying preferences for academic fields. We further demonstrate the plausibility of this conclusion by means of theoretical analysis of a model of student choices of academic concentrations. An important takeaway from our results is that it is one's weaker underlying preference for a field of study that is likely to make a student more "sensitive" to grades received in it, rather than the other way around. This is contrary to a commonly suggested understanding that it is the underlying stronger sensitivity to grades that makes students possessing such characteristic less attached to "tougher" grading fields.

JEL codes: I23, I24, J24, D21.

---

*Indiana University; CESifo
†Indiana University
‡Indiana University

[1] Link to most recent version of this paper: `https://drive.google.com/file/d/1CITkz5iXHa_DSYoUdlPwif2cdMzeh69H/view`

[2] The acronym derived from the names of the constituent groups of disciplines: Sciences, Technology, Engineering, and Mathematics.

1 Introduction

A new direction in the economics of higher education, which developed over the last two decades, has focused the attention on students' choices among the fields of study. This is well justified by the evidence that the choice of college major is becoming a stronger determinant of the variation in career earnings than the choice between going to college or not. In other words, the variation of college *major* premia is overtaking the average college premium (see James, 2012, Hershbein and Kearney, 2014). Furthermore, Altonji et al. (2015) and Kirkeboen et al. (2016) affirm that major choices are dominant determinants of future earnings even controlling for the quality of peers and the higher education institution, and that in fact the effect on earnings from attending a more selective institution is dominated by the payoffs to a field of study. Accordingly, the literature also provides strong evidence that students' expectation of future earnings associated with college majors is a significant positive determinant of their decisions to choose among them: see Berger (1988), Montmarquette et al. (2002), and Arcidiacono (2004, 2012) among others.

The facts of substantial of gender differences in student choices of academic disciplines are well known, among which the significant underrepresentation of women in STEM and their overrepresentation in Education are the most striking. For instance, Gemici and Wiswall (2014) document persistence of a significant gender gap in major choices in STEM and Business in favor of men, as well as a gap in the opposite direction in the Social Sciences (excluding Economics) and Humanities (SSH). The former is especially striking given that women increasingly overtake men in terms of the overall college enrollments, as well documented by Goldin et al. (2006)[3]. Given the aforementioned differences in college premia across fields, with STEM leading in lucrativity, such gender imbalances have obvious implications for the distributions of career earnings of college educated men and women, particularly the wage gap. Among potential explanations for the gender gap in the pursuit of lucrative fields in college, some authors, e.g., Altonji et al. (2015), Zafar (2013), Rendall and Rendall (2014), raise the conjecture that there are differences between genders in the importance students attach to pecuniary and non-pecuniary benefits associated with future occupations.

In addition to gender differences in propensity to enroll in STEM and other quantitatively oriented majors, higher rates of attrition from such majors further contribute to the gender imbalances in the gender compositions of the total numbers of degrees awarded in academic disciplines. Significantly higher attrition rates from STEM among women are particularly well documented: see, e.g., Chen and Soldner (NCES, 2013).

A substantial literature focuses on the influence of grades students receive in courses on their decisions to persist in the corresponding disciplines. The studies show, in particular, that students do indeed condition their choice of a more lucrative major on the perceived probability of their success in it, expressed in grades, which they expect to receive in this major along with the effort this will entail[4] – see, e.g., Montmarquette et al. (2002) as well as the findings in Rask (2010), and Stinebrickner and Stinebrickner (2014) that the trade-off between the relatively high expected future wages associated with STEM disciplines and the grades earned there, is a key determinant in students' decisions whether to persist in these majors. Additionally, Arcidiacono (2004) finds that although expectation of a lucrative career is a significant factor in student choices of a major, poor performance causes one to switch to a less lucrative major.

Indeed, as a student makes an initial choice of a major, the grades earned along the way signal (to the student as well as, potentially, to future employers) the need to update his/her expectation of success in said major and may affect his/her decision as to whether to persist in it or to switch to an alternative major. Further, it is well-recognized that grades vary strongly across disciplines (lower in STEM, higher in Humanities and Social Sciences, save for Economics) and that they do have an effect on student choices of majors – see a survey by Aachen and Courant (2009). This overall trade-off between expected future wage earnings and the academic standards (and accordingly, the study effort) required in the corresponding majors can be well understood in the context of labor market equilibrium, as higher expected returns of an occupation are potentially balanced out, for a marginal candidate, by the higher effort cost of qualifying for it.

[3]A recent literature explores whether a difference in gender-specific college premia may help explain the gender gap in the overall college enrollment (Becker et al., 2010, Ashworth and Ransom, 2019).

[4]See also Babcock and Marx (2010) for comprehensive evidence that the downward trends in study effort, measured by time devoted to study in college classes, as an essential choice variable motivating student decisions in college.

The above facts lead some scholars to conjecture that women tend to exhibit relatively stronger responsiveness (or "sensitivity") to grade signals than men, and that this is a factor in higher rates of attrition among women from disciplines such as STEM characterized by their less generous grading standards. Montmarquette et al. (2002), who find a strong correlation between perceived ability and the choice of lucrative major, show that this correlation is weaker among women; specifically, controlling for grades, women are less likely to be attached to lucrative majors than men. Put differently, according to Montmarquette et al. (2002), women exhibit stronger sensitivity to grades they receive, hence weaker attachment to lucrative majors. Rask and Tiefenthaler (2008) find strong evidence for this gender gap in grade sensitivity in students' decisions whether to persist in their study of Economics, while Ost (2010) reaches a similar conclusion when it comes to Physics. Feng et al. (2018), who also use the Indiana University data, document stronger responsiveness of women to grades received in STEM classes (particularly in relation to grades earned in other disciplines) in terms of decisions against persisting in STEM. Note also the observation by Chen and Soldner (2013), who use NCES survey data, of greater aversion on the part of women and minorities to an excessively "competitive climate" in STEM – an increasingly common claim in the literature.

To sum up, there are two lines of economics reasoning offering competing explanations for gender gap in the choices of majors. A large share of the literature referenced immediately above, which is focused on student responses to grades, advances a *behavioral* explanation rooted in gender differences in reaction to grades: it suggests that women generally exhibit stronger sensitivity to inferior grade performance; this leads to their relative aversion to and attrition from the disciplines, such as STEM and Business/Economics, which tend to assign relatively low grades. The second line of reasoning (see, e.g., Altonji et al., Zafar, Rendall and Rendall, *op. cit.*) finds evidence in support of a claim that women place lower weight on pecuniary benefits (and greater weight on the non-pecuniary ones) associated with a career than do men in their choices of college majors (which may, arguably, be rooted in economic incentives based on present differences in marriage markets and family roles in the society). From this, one can then derive what can be called a *"homo economicus"* explanation for gender differences in sensitivity to grades. Since more lucrative college majors do tend to assign relatively lower grades, the individuals who are relatively less invested in the lucrativity of future careers, will be less inclined, *ceteris paribus*, to accept the disutility of the grade and effort downside of such choices. It then follows that women would be relatively more averse to the aforementioned majors offering the trade-off of more lucrative pay for poorer grades (as well as inferior other non-pecuniary benefits).

A logical approach to examine the two competing explanations is to investigate whether the relatively stronger sensitivity to grades women exhibit in STEM and Business/Economics are sustained in less lucrative academic disciplines such as Humanities and Social Sciences. Such analysis has not been heretofore performed in the literature. The present paper pursues exactly this task. A comprehensive Indiana University-Bloomington Learning Analytics dataset allows us to look at student responses to grades, in terms of persistence in an initially chosen academic field, in a broader context across the entire spectrum of academic disciplines. The same advantage allows us to not only assess students' persistence in or transition from an academic category, such as STEM, but to differentiate student transitions with respect to the potential destination categories, such as Business and Economics, Social Sciences and Humanities, etc., depending on the grade signals they receive from their performance in the classes taken in these alternative (destination) academic categories.

This comprehensive characterization of student persistence throughout the academic categories available at the university, along with cross-category transitions is novel and offers a broader outlook on gender differences in field preference and persistence patterns, and in the effects of grade performance on these decisions.

The key results of our analysis reveal that 'grade sensitivity' of a student is not a gender-specific behavioral attribute *per se*. Although women do indeed tend to exhibit it more than men in their patterns of migration out of STEM as well as Business and Economics, this pattern does not emerge in students' migrations from Social Sciences (excluding Economics) and Humanities, where women demonstrate either equal or relatively lower responsiveness to grades received. This suggests that grade sensitivity reflects students' preferential attachments to academic disciplines, akin to the obvious fact that consumers' responsiveness to prices depends on their utility valuations of the corresponding products. This understanding supports the aforementioned *"homo economicus"* conjecture based on gender differences in the valuations of pecuniary and non-pecuniary benefits of careers associated with academic disciplines. It also presents an argument that it is the underlying weaker preferential attachment to a field that makes a student more sensitive to grades received in it, not the other way around. In Section 5, we offer a theoretical framework for the above reasoning.

The paper is structured as follows: Section 2 describes the data used in our empirical analysis. Section 3 outlines the econometric model. Section 4 states the results, and presents robustness checks for our results. Section 5 offers a theoretical outline of the decision-making process students go through in selecting a field of study, and subsequently persisting in it or switching to an alternative. Section 6 concludes.

2 The Data

The data set used for this paper contains selected demographic and academic information for 50,545 Indiana University Bloomington domestic[5] undergraduate students. This student data comes from Indiana University's Learning Analytics data set, which contains detailed information for every undergraduate student enrolled at IU from Fall 2006 up through Spring 2017. The Learning Analytics data set not only includes a range of demographic variables for each student, but also records each student's semester-to-semester academic activity. For example, it tracks every class taken by a student, every grade received, and lists all majors declared during the student's time at IU.

The focus of this paper is on students' academic decisions as they transition from their second to third year at IU. A spotlight on a specific timeframe of students' decisions is essential for clean analysis. A decision to switch one's academic field at the end of year one as opposed to such decision in a later year entails different opportunity costs and is based on different amounts of information regarding his/her academic interests, ability, and understanding of academic options. Our particular focus on the transition from the second to third year of studies is well justified by the fact that IU students typically finalize their selection of major at the start of the third year, by which time they have typically taken sufficient number of prerequisite courses allowing for a more informed decision based on their relevant performance. By the same token, by this point in their studies students face non-trivial opportunity costs of switching to a different academic category with a different set of academic prerequisites[6].

Our focus on students transitioning from their second to third year necessitates some trimming of the population in the Indiana University's 2006-2017 Learning Analytics data set. Specifically, we exclude students from cohorts beyond the Spring 2015 since the data set does not allow us to observe their third-year action. For the same reason, we must also exclude students who graduate at the end of their second year (likely a result of having transferred sufficient credits upon entry at IU).

Thus, we are interested in student choices when transitioning into the third year of their studies at IU whether to persist in their present academic discipline or to switch to an alternative field. We examine, in particular, whether potential male-female differences in grade sensitivity are reflected in these decisions in transition from the second to third year of studies. Before we can analyze the impact of each gender's grade sensitivity on academic path persistence and migration, we first clarify our definition of 'academic path,' as well as the type of action, which constitutes a change in academic path.

An academic path, for the purposes of our analysis, is defined by two points: the academic category to which a student belongs in his/her second year, and the academic category they find themselves in at the start of the third year. The term 'academic category' refers to a grouping of majors which have similar

[5]We have excluded IU Bloomington's the foreign students, because this population features systematic distinctions from the domestic one, in terms of the cost of education and, due at least in part to a related selection bias, in preferences for academic fields. Additionally, international students were not required to report SAT scores - an important control variable used in our regression. Also important that foreign students are likely substantially different from domestic students in other ways: for instance, the total cost of being at IU is substantially higher for them; the weights they place on pecuniary career benefits may also be different. Given these unknowns it is reasonable to expect that including foreign students might present such issues of unobserved heterogeneity.

[6]The focus on decisions at the half-way point in a typical duration of college studies is natural and common in the literature. See, for instance, Arcidiacono (2004) whose analysis of NLS72 data is focused on students' transitions from the initial academic disciplines two years after the start of their studies.

academic foundations and tools. All majors were grouped into the following five academic categories:[7] STEM, Business and Economics (BE), Social Sciences and Humanities (SSH), Other Professional (OP), and Education (Educ). The criterion of this method of aggregation of university majors is academic similarity; in other words, grouping likes with likes. The more mathematical and/or empirically analytical majors (STEM and BE) are separated from the majors that focus on reading, writing and verbal communication skills (SSH), and from the majors that focus on developing professional career-specific skills (OP and Educ). This academic categorization also corresponds well with the breakdown of majors in terms of career earnings. Furthermore, it is consistent with the existing literature and allows for easier comparison of results.

Since our analysis focuses on transitions between broad academic categories, it does not study student migration across majors which belong to the same academic category. For example, students who change their major from Chemistry to Biology remain in the STEM academic category, and are seen here as having persisted on the STEM path.

We note that transitions across broad academic categories as defined here (e.g., from Chemistry in STEM to English, which belongs to SSH) are substantially more costly, in terms of time and effort, than transitions within a category (e.g., from Chemistry to Biology, both in STEM) because the former entails fulfilling a new set of academic prerequisites, i.e., additional coursework, even before qualifying for taking courses in the new major *per se*. Accordingly, such transitions are also characterized by acquiring qualitatively different skill sets. This ensures the focus on the most fundamental changes in students' academic trajectories, which are also the most consequential in subsequent career choices.

It is also important to note that our focus on second to third year academic transitions, although best suited for the study of cross-category migrations, misses much of the action when it comes to the decision to drop out of the university. Most of the latter type of action (54%), according to our data set, occurs during or at the end of the first year of studies, with just over 24% of dropout happening during or at the conclusion of the second year of studies at IU[8].

The following Table 1 helps establish a general sense of student attrition by academic category aggregating students' outflows from and inflows into each category throughout their studies at IU (i.e., not just at the end of the second year). To this end, Table 1 aggregates all academic category related activities for students in the Fall 2006 through Fall 2010 cohorts. By excluding students associated with cohorts beyond Fall 2010 we ensure that the *entire* IU careers of remaining students (from first to final academic action at IU) are observed in the data set.

Table 1: Graduating Shares by Academic Categories in 2006-2010 Cohorts

Academic Category	Starting Size		Proportion		Received BA/BS		Proportion		Dropped Out		Proportion	
	Female	Male	Female %	Male %	Female	Male	Female %	Male %	Female	Male	Female %	Male %
STEM	2173	2233	12%	14%	1622	2003	12%	17%	400	500	12%	14%
BE	3093	6762	17%	42%	2058	4673	15%	39%	402	1260	12%	35%
SSH	8425	5789	47%	36%	5481	3521	41%	29%	1586	1476	48%	41%
OP	3091	1091	17%	7%	3894	1824	29%	15%	709	278	22%	8%
EDUC	1140	215	6%	1%	435	85	3%	1%	197	63	6%	2%
Total	17922	16090	100%	100%	13490	12106	100%	100%	3294	3577	100%	100%

* Includes all students in these cohorts graduating in a given category, including those in it from the start and those ultimately migrating to it from other categories.

Two remarkable takeaways from Table 1 is a fairly dramatic shrinkage in Social Science and Humanities (SSH) share in student choices by the time of graduation, and an even more significant rise in the share of Other Professional (OP) disciplines, which prove to be an attractive *ultimate* destination category despite an apparent relatively low initial appeal.

[7]STEM (Science, Technology, Engineering, Mathematics) categorization is based off the Homeland Security definition and categorization of STEM majors; the OP ("Other Professional") category includes Indiana University School of Informatics (except for its constituent majors which the government categorizes as STEM, as stated above), School of Public and Environmental Affairs (SPEA), School of Public Health, School of Social Work, Apparel Merchandising and Interior Design, and a few others.

[8]See the Timing of Dropout chart in the Appendix.

Before returning to the main data set of 50,545 students specifically built to examine second to third year student academic transitions, it is important to first determine whether this timeframe focus may miss a substantial share of transitions. Table 2 below, showing the distribution of students in the data set in terms of the number of academic categories to which they belong during their academic careers at Indiana University, largely alleviates such concern.

Table 2: Number of Academic Categories with which Students are Associated

	1 Category	2 Categories	3 Categories	4 Categories	Total
Student Nbr	32,425	16,059	2,004	57	50,545
Percentage	64.2%	31.8%	4.0%	$< 1\%$	100%

We see that the majority of students (64.2%) are only ever associated with a single academic category, i.e., persist in this field throughout their entire academic career. However, there is a significant share of students in this population, 31.8%, who do switch their primary majors once across categories, such that they become associated with a second academic category. Very few students (approximately 4%) are ever part of three or more academic categories.

Returning to our main focus on the decisions of the 50,545 students in the data set made between the second and third year, the following Table 3 gives a more detailed look at the type of action taken. This table shows the number of students who started in a given academic category during their second year, and the percentage of those students who are female is shown in parentheses. Going across a row, one can see how many students remained in their original category as they transitioned into their third year, as well as how many students migrated to other categories. The bottom row shows how many students start their third year in a given category.

Table 3 also offers preliminary evidence of substantial imbalances between men and women in certain academic categories. For example, although women represent 51% of all students at IU, they account for less than one-third of the Business and Economics students in both the second and third year. Women are also underrepresented in STEM at both the start and end of our analysis (only 44% and 43% during the second and third years respectively). There is also a notable underrepresentation of men in the SSH, OP, and especially in the Education category.

Table 3: Persistence and Exits across Academic Categories

Second Year		**Third Year Actions**					
Category	**Starting Size**	**STEM**	**Bus & Econ**	**SS & Hum**	**Other Prof**	**Educ**	**Dropout**
STEM	8,025 (44%)	**6,784 (43%)**	116 (29%)	432 (54%)	345 (63%)	12 (67%)	336 (34%)
Bus & Econ	13,984 (29%)	416 (19%)	**11,982 (30%)**	526 (36%)	637 (33%)	16 (63%)	407 (20%)
SS & Hum	16,999 (59%)	637 (42%)	451 (31%)	**13,597 (61%)**	1,226 (65%)	146 (73%)	942 (8%)
Other Prof	9,265 (69%)	149 (60%)	169 (30%)	364 (65%)	**8,106 (70%)**	50 (80%)	103 (67%)
Educ	2,172 (85%)	14 (50%)	8 (38%)	91 (71%)	90 (86%)	**1,866 (87%)**	112 (80%)
Total	50,545 (51%)	8,000 (42%)	12,726 (30%)	15,010 (60%)	10,404 (67%)	2,090 (86%)	2,315 (46%)

Table 3 further demonstrates that each academic category has significant numbers of students migrating in and out of it. In other words, the dataset exhibits substantial cross-category migration, despite the fact that it aggregates cross-major migrations within each of these broad categories. For example, STEM loses 432 students to SSH, of which 54% are women, and it loses another 345 students to the Other Professional (OP) category, of which 63% are women.

The descriptive statistics collected in the following Table 4 offer a general picture of gender differences in grade distributions across the academic categories. Student information is displayed across five GPA bins, which descend from the maximum 4.0 GPA in intervals of 0.5 until the GPA of 2.0, which corresponds to the letter grade C, with students possessing GPA of 2.0 or lower, i.e., those struggling to meet the minimum requirements in their field, are grouped together in the lowest GPA bin.[9]

Two main if striking takeaways from these statistics are (a) that women exhibit remarkably superior average grade performance in all academic categories and (b) that a large part of this distinction results from increasing gender sorting in lower grade intervals. In particular, the gender imbalance in favor of men in STEM and BE gets substantially stronger as one moves down the grade scale.

Table 4: Grade Distributions Across Academic Categories by Gender

Average GPA across Categories

Category	Female	Male
STEM	2.94	2.83
BE	3.26	3.10
SSH	3.28	3.06
OP	3.47	3.27
EDCU	3.63	3.41

Soc. Sci. & Hum. Population

SSH Grade	Female	Female %	Male	Male %
A+ to A-	4133	41%	2031	29%
B+ to B-	3368	33%	2051	30%
B to B-	1700	17%	1525	22%
C+ to C	581	6%	828	12%
C- & Below	306	3%	475	7%
Total	10,088	100%	6,910	100%

STEM Population

STEM Grade	Female	Female %	Male	Male %
A+ to A-	928	26%	1048	23%
B+ to B-	938	27%	1101	24%
B to B-	767	22%	991	22%
C+ to C	476	14%	702	16%
C- & Below	402	11%	672	15%
Total	3,511	100%	4,514	100%

Other Prof. Population

OP Grade	Female	Female %	Male	Male %
A+ to A-	3799	59%	1187	41%
B+ to B-	1829	28%	1019	35%
B to B-	557	9%	473	16%
C+ to C	171	3%	148	5%
C- & Below	94	1%	88	3%
Total	6,450	100%	2,915	100%

Bus. & Econ. Population

BE Grade	Female	Female %	Male	Male %
A+ to A-	1582	39%	2895	29%
B+ to B-	1558	38%	3548	36%
B to B-	631	15%	1947	20%
C+ to C	209	5%	900	9%
C- & Below	124	3%	590	6%
Total	4,104	100%	9,880	100%

Education Population

EDUC Grade	Female	Female %	Male	Male %
A+ to A-	1417	76%	184	58%
B+ to B-	320	17%	78	25%
B to B-	67	4%	30	10%
C+ to C	33	2%	8	3%
C- & Below	20	1%	15	5%
Total	1,857	100%	315	100%

We shall now briefly discuss the definition of *dropout* from the University this paper uses when cataloguing student choices and academic activities. The Learning Analytics data set does not have an identifier for students who drop out, since it is always possible that a student who stopped enrolling in classes without having completed a degree, may choose to resume studies at the university at a later time. However, upon examination of the data, we found that the cases of students resuming studies at IU if they are absent from enrollment rows for more than one year, are very rare. The following Table 5 shows the distribution (in percentage terms) of students who take breaks from studies at IU in terms of the lengths of the corresponding intervals.

[9] Standard conversion of letter grades into numerical GPA scale is as follows: A = 4.0; A- = 3.7; B+ = 3.3; B = 3.0; B- = 2.7; C+ = 2.3; C = 2.0; C- = 1.7, etc.

Table 5: An Examination of Student Break Length

Length of Break	Number of Students	Percentage
No Breaks	72,125	79.1%
One Semester	12,104	13.3%
1 year	3,017	3.3%
1.5 years	1,400	1.5%
2 years	961	1.1%
2.5 years	534	0.6%
3 years	387	0.4%
More than 3 years	640	1%
Total Number of Students	91,168	100%

We see that the majority of students, 79.1%, do not take any time off from school. There is a small percentage of students who do take a one-semester break before resuming classes (13.3%), and an even smaller percentage of students who take a one-year break (3.3%). Students who take one and half or more years off from school cumulatively account for 4.6% of the student body. The break-length-trends seen in this table suggest that it is very unusual for students to stop taking classes for one and half years, and then resume working towards an IU degree. Thus, we identify the student as dropout if he/she has not yet earned a degree and has not enrolled in any classes during the last one and half years of our data set. In other words, students are categorized here as dropouts if they have stopped working towards a degree of any kind as evidenced by their lack of enrollment for at least a year and a half. Recall that our data set starts in Fall 2006 and ends in Spring 2017. Therefore, students enrolled in the Fall 2015 are the last cohort among whom we can identify the dropouts.

3 Econometric Model

We model an individual student's decision problem at the beginning of the third year as follows. We model student's choice of academic categories, i.e., according to our dataset, the choice from the following category set: STEM=1, BE=2, SSH=3, OP=4, EDUC=5, DROP=6. Given her academic starting category k during the second year, student i needs to determine her field j in the coming year. This corresponds to a decision as to whether she should persist in her current field k, i.e., setting $j = k$, or to migrate to one of the other four academic categories, i.e., $j = 1, \dots, 5$, with $j \neq k$, or drop out, i.e., selecting $j = 6$. Individual student's indirect utility of choosing action category j, denoted by $U_{ij|k}$, is expressed as follows:

$$\mathrm{U}_{ij|k} = X(i)\beta_{j|k} + \epsilon_{ij|k},$$

where $\beta_{j|k}$ captures how student's characteristics affect their tastes toward different academic categories, $\epsilon_{ij|k}$ represents the idiosyncratic taste shocks, and $X(i)$ is a vector of student i characteristics given by demographic, socio-economic, and academic variables.

In particular, the demographic variables include students' gender and ethnicity. The role of students' gender in their academic choices is a key variable in our analysis. As noted earlier, it is widely observed that women are underrepresented in STEM, Business and Economics – the "lucrative" fields, i.e., those associated with relatively high post-graduation earnings. By exploring the gender differences in students' persistence in STEM in particular, we are able to infer how amenable each gender is to continuing on or pursuing academic options outside of STEM, thus allowing us to characterize male versus female attachment to STEM. We then go on to characterize male versus female persistence in other academic categories to see whether and how gender differences in sensitivity to grades differ from our findings for the STEM category.

The socioeconomic variables allow us to control for financial and educational differences in students' family backgrounds. In particular, *Pell grant eligibility* is a binary variable, which equals 1 if the student is eligible for a Pell grant, and zero otherwise. Pell grants are awarded based on their families' low-income status, acting as a rough indicator of a student's family's financial circumstances and its ability to provide the student with financial support. *Residency* indicates whether student a qualifies for in-state tuition as a resident of Indiana. Since in-state resident students are charged dramatically lower tuition than non-residents, including the students' residency status in the model allows us to control for differences in students' opportunity costs. *First-Generation Student* indicates a student whose parents had no education beyond the secondary level. This information is helpful to control any large differences in the educational background of a student's family, which is known to affect students' performance and choices through a number of factors, such as information about college preparation, resources and life – both academic and social.

We use student's GPA at the point of matriculation at IU and their $Math$ and *Verbal* SAT scores[10], i.e., the pre-college academic aptitude measures, to proxy students' academic ability and the level of college preparedness. We also include two variables to capture students' academic history in the first two years: The variable *enter as intended* indicates whether or not a student was admitted directly into the school or program, to which he/she originally applied; the variable *switched before year two* indicates whether or not the student had changed his/her academic categories before the start of year two.

A key variable of interest is the student's year-two category GPA, which allows us to investigate how students respond to their academic performances in the current category when they consider whether to persist in it or to migrate. Note that this decision process also depends on students' performance in alternative categories as students might shop around for other academic fields and make a decision, in part, based on their performance in these other fields. Thus, we also include students' other category GPA information in the model. Since not all students have taken courses outside of their current academic categories, some would have no information regarding their performance in other fields. To control for this lack of information situation, we include four *lacking-information dummy* variables associated with the *other category GPAs*, which equals to 0 if the student has taken courses in that field, and 1 otherwise.

Given a student's academic starting point and her characteristics, student i chooses her academic path to maximize her indirect utility, as expressed below:

$$\mathrm{U}_{ij^*|k} = max_{j=1,\ldots,6} U_{ij|k}$$

Due to identification issues, we normalize the deterministic part of the indirect utility which a student derives from dropping out as zero, i.e., $\beta_{6|k} = 0$. If the taste shocks are from a type I extreme value distribution and are independent and identically distributed across categories, we can characterize a student's academic decision using the multinomial logistic expression. That is, conditional on the student's starting point, k, the probability of student i choosing action j in their third year, denoted as $\pi_{ij|k}$, is given by

$$\pi_{ij|k} = \frac{e^{X(i)\beta_{j|k}}}{1+\sum_{n=1}^{5} e^{X(i)\beta_{n|k}}}$$

which is exactly a multinomial logistic regression and can be easily estimated. Note that the coefficients are identified up to scale so that only the signs of the coefficients (and not their absolute values) carry meaning by indicating the direction of a variable's effect. Thus, we focus on the marginal effect, which, for a binary explanatory variable such as gender, represents the difference in the associated probability.

4 Results

The student population in the data set is partitioned according to students' starting academic categories (STEM, BE, SSH, OP, and EDUC). We estimate the multinomial logistic regression model and evaluate the gender effect for each of the five subpopulations separately. Even though the current category GPA enters the regression model as a continuous variable, we aggregate the gender effect in academic category decisions for the aforementioned GPA bins.

The estimated transition probabilities obtained for each academic category/GPA bin/gender combination indicate the likelihood of persisting in or switching to an alternative academic category (whereby we included dropping out in the list of options), conditional on the student's second year academic category. The following

[10] The SAT scores assigned to each student in the data set are either the ones the student submitted with his/her application to IU, or the scores obtained by a standard conversion procedure from the student's ACT scores, the alternative aptitude test, which meets IU application requirements.

Table 6 presents the resulting estimates for select starting academic categories: STEM, BE, and SSH (see Table A in the appendix for the full set of results).

The columns, which report female and male persistence for each of the selected academic categories, reveal two noteworthy trends. First, the men in the STEM and BE academic categories (the top and the center panels of Table 6, respectively) consistently display higher estimated probabilities of persistence relative to their female counterparts. Furthermore, this gender differential in persistence grows as the grade received declines.

Table 6: **Female & Male Cross-Category Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	93.0%	95.4%	1.3%	1.5%	2.5%	1.6%	2.7%	1.1%	0.0%^{X}	0.0%^{X}	0.5%	0.4%
B+ to B	88.9%	93.2%	1.3%	1.5%	3.9%	2.5%	4.7%	1.9%	0.1%^{X}	0.0%^{X}	1.0%	0.9%
B to B-	82.8%	89.3%	1.2%	1.5%	6.3%	4.1%	7.1%	2.9%	0.2%*	0.1%^{X}	2.5%	2.1%
C+ to C	73.2%	82.1%	1.3%	1.6%	9.3%	6.4%	9.7%	4.2%	0.5%**	0.1%^{X}	6.1%	5.6%
C - and below	51.5%	61.9%	1.1%	1.5%	15.9%	12.1%	11.8%	5.5%	0.8%*	0.3%^{X}	18.9%	18.8%

Business & Economics-starting Population

BE Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	1.0%	1.2%	95.0%	96.5%	1.5%	0.8%	1.7%	1.0%	0.2%*	0.0%^{X}	0.6%	0.6%
B+ to B	2.1%	2.6%	87.5%	90.9%	4.0%	2.1%	4.5%	2.8%	0.2%**	0.0%*	1.6%	1.5%
B to B-	3.5%	4.4%	75.3%	81.4%	7.8%	4.4%	9.4%	6.1%	0.4%**	0.1%*	3.6%	3.6%
C+ to C	5.0%	6.8%	58.4%	66.7%	13.0%	7.9%	15.3%	10.4%	0.6%*	0.1%*	7.8%	8.1%
C - and below	6.2%	9.0%	37.4%	45.5%	20.7%	13.7%	20.1%	14.8%	0.4%^{X}	0.1%^{X}	15.2%	16.9%

Social Sciences & Humanities-starting Population

SSH Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	2.6%	4.4%	1.3%	3.3%	89.5%	87.4%	4.5%	3.4%	0.7%	0.4%	1.3%	1.2%
B+ to B	3.0%	5.0%	1.7%	4.3%	83.3%	81.5%	8.2%	6.1%	1.0%	0.5%	2.8%	2.5%
B to B-	3.2%	5.4%	1.8%	4.6%	76.3%	75.2%	11.0%	8.3%	1.3%	0.6%	6.4%	5.9%
C+ to C	3.0%	5.3%	1.7%	4.3%	64.2%	64.4%	13.5%	10.4%	2.0%	1.0%	15.6%	14.6%
C - and below	2.4%	4.3%	1.2%	3.3%	40.6%	41.6%	12.7%	10.0%	2.2%	1.1%	40.9%	39.6%

X **not significant at 5% level, * 5% level, ** 1% level, all other results are significant at the 0.1% level.**

The second noteworthy trend in question can be seen in the SSH columns (which show probabilities of persistence) in the bottom panel of Table 6. These SSH persistence columns show a partial reversal of the gender persistence differential seen in the STEM and BE panels. Specifically, women display higher levels of persistence relative to their male counterparts at higher grade levels, and this differential diminishes up to parity as the earned grade declines, until GPA levels drops to 'C- & below' where men become the somewhat more persistent gender.

Based on the comparisons of the estimated persistence and transition probabilities of male and female students in our regression model seen in Table 6, we now proceed to analyze the statistical significance of the gender differences seen between the corresponding persistence and transition probabilities. The results of this analysis are displayed below in Table 7.

The column entries in Table 7 correspond to the female-male column-pairs of Table 6 and display the statistical significance of the differences in corresponding male and female persistence or transition probabilities (specifically, probability for a male student in a group in question, minus the corresponding probability for a female student).

Table 7: **Gender Differences in Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	2.48%*** (0.00461)	0.23% (0.00292)	-0.936%*** (0.00229)	-1.65%*** (0.00260)	-0.04% (0.000281)	-0.08% (0.000611)
B+ to B	4.26%*** (0.00618)	0.25% (0.00287)	-1.40%*** (0.00340)	-2.85%*** (0.00402)	-0.09% (0.000625)	-0.17% (0.00136)
B to B-	6.49%*** (0.00870)	0.28% (0.00274)	-2.13%*** (0.00531)	-4.15%*** (0.00570)	-0.17% (0.000991)	-0.33% (0.00313)
C+ to C	8.89%*** (0.0125)	0.36% (0.00293)	-2.85%*** (0.00773)	-5.52%*** (0.00797)	-0.32% (0.00197)	-0.56% (0.00722)
C- & below	10.40%*** (0.0168)	0.41% (0.00266)	-3.85%** (0.0128)	-6.26%*** (0.0106)	-0.54% (0.00387)	-0.16% (0.0168)

Business & Economics-starting Population

BE Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	0.18% (0.00136)	1.43%*** (0.00269)	-0.74%*** (0.00134)	-0.68%*** (0.00135)	-0.14%* (0.000679)	-0.06% (0.000827)
B+ to B	0.47% (0.00285)	3.42%*** (0.00598)	-1.84%*** (0.00328)	-1.77%*** (0.00354)	-0.18%* (0.000810)	-0.10% (0.00218)
B to B-	0.99%* (0.00463)	6.11%*** (0.0102)	-3.38%*** (0.00631)	-3.36%*** (0.00706)	-0.31%* (0.00148)	-0.04% (0.00469)
C+ to C	1.78%** (0.00669)	8.30%*** (0.0137)	-5.13%*** (0.0103)	-4.86%*** (0.0112)	-0.45% (0.00270)	0.34% (0.00928)
C- & below	2.78%** (0.00851)	8.16%*** (0.0134)	-7.03%*** (0.0158)	-5.30%*** (0.0147)	-0.34% (0.00279)	1.73% (0.0157)

Social Sciences & Humanities-starting Population

SSH Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	1.78%*** (0.00315)	1.98%*** (0.00255)	-2.10%*** (0.00469)	-1.17%*** (0.00238)	-0.39%*** (0.000990)	-0.11% (0.000972)
B+ to B	2.06%*** (0.00342)	2.59%*** (0.00298)	-1.82%** (0.00621)	-2.08%*** (0.00420)	-0.52%*** (0.00128)	-0.24% (0.00204)
B to B-	2.23%*** (0.00353)	2.76%*** (0.00300)	-1.10% (0.00784)	-2.71%*** (0.00563)	-0.67%*** (0.00178)	-0.51% (0.00450)
C+ to C	2.23%*** (0.00341)	2.64%*** (0.00297)	0.16% (0.0106)	-3.09%*** (0.00697)	-1.00%*** (0.00300)	-0.94% (0.00956)
C- & below	1.86%*** (0.00306)	2.05%*** (0.00289)	1.09% (0.0123)	-2.65%*** (0.00726)	-1.10%** (0.00397)	-1.25% (0.0157)

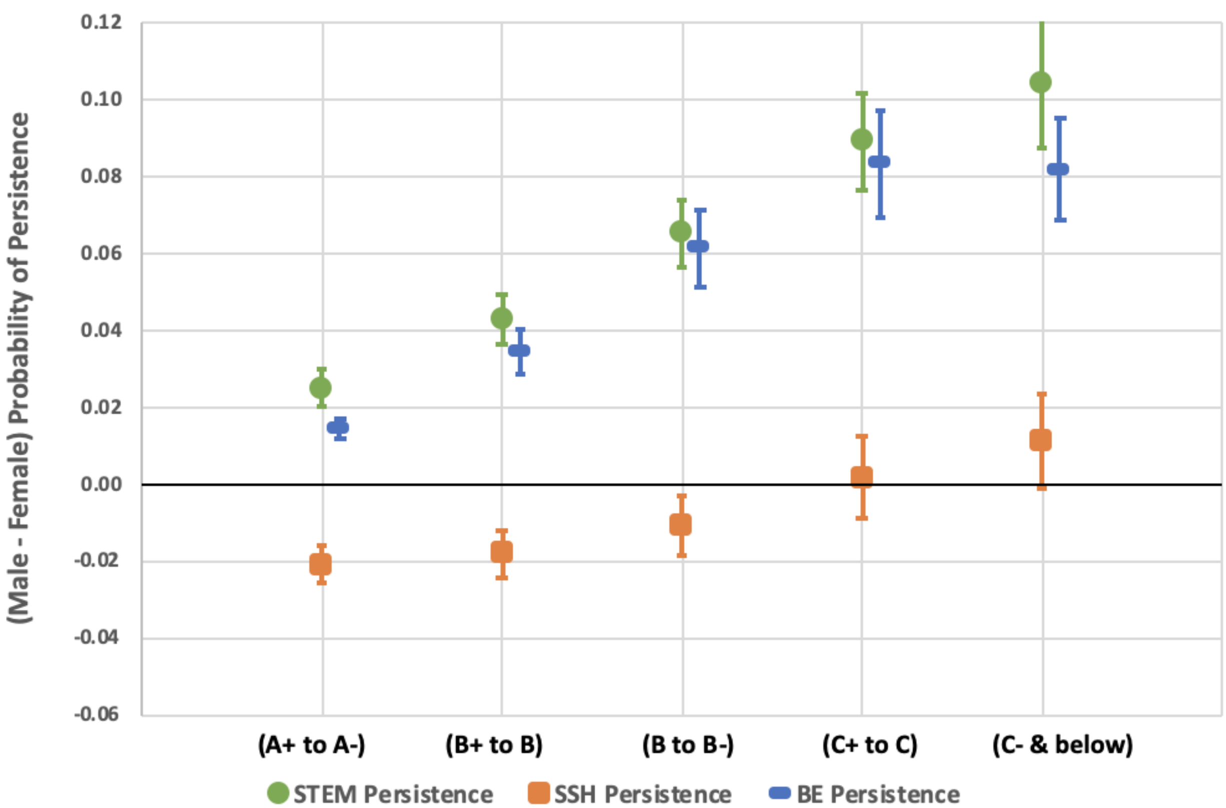
Standard errors in parentheses
***** $p < 0.05$, ****** $p < 0.01$, ******* $p < 0.001$

Since female student outcomes constitute the benchmarks for this table of marginal effects, a positive value indicates that there are relatively more men than women in that particular persistence/migration group. Accordingly, any negative numbers indicate that there are relatively more women than men in that particular group. For example, the value of 8.9% for 'C+ to C-' students who started their second year as STEM students indicates that, upon earning a 'C+ to C' grade in STEM, the estimated probability of persisting in STEM is 8.9 percentage points higher if the student is male rather than female. Similarly, looking at the same grade range of 'C+ to C' in the STEM table, but now at the SSH column, the negative entry of -2.9% implies that male 'C+ to C' grade-level STEM students are less likely to migrate to SSH than their female counterparts.

We now turn to the statistical significance of gender differentials in persistence seen in the selected academic categories[11]. We start with persistence in STEM reported in the first column in the top panel of Table 7. The fact that *all* the values in this column are positive indicates that male persistence in STEM exceeds female persistence across all STEM grade levels. In other words, across all STEM grades earned and controlling for other regressors, being male increases the probability that a student will persist in a STEM discipline. The second trend we noted earlier in this STEM column is that the values *increase* as one moves from the top to the bottom of the column. This implies that as the grade received in STEM declines the stated gender imbalance widens such that the women in the corresponding grade bin choose to exit the field at a relatively higher pace than their male counterparts. This results in increasing under-representation of women in STEM as one moves down the grading scale. This growing gender differential implies that women are relatively more responsive to the grades they receive in STEM, further indicating that earned grades play a relatively larger role in female students' decision to abandon this academic category.[12]

The statistical significance of the gender imbalance in STEM persistence is reported in Figure 1, represented by the circles. This figure plots the values from the columns corresponding to persistence in a starting category in the corresponding panels of Table 7 and adds standard error bars to each point, which helps determine whether the difference between the values are statistically significant. If the standard error bars do not overlap, then the two points are statistically different from one another. Although not every point is statistically different from the next, the overall trend for persistence in STEM does appear to be significant – that *male persistence increasingly dominates female persistence as the grade received in STEM declines.*

Figure 1: Gender Differences in Persistence Probabilities



The gender differential in student persistence in the Business and Economics category (BE) parallels the STEM trends discussed above. Indeed, the positive and increasingly large values seen in the BE column

[11] See the appendix Table B for gender marginal effect estimates for the *full* set of academic categories

[12] It is noteworthy that this phenomenon also likely contributes to the ultimate superior grade performance of women over men among students who persist in STEM.

of the Business Economics-starting Population panel (i.e., the second column of the second panel of Table 7) not only indicate *that men are consistently more persistent than their female counterparts,* but also that this male tendency to out-persist women increases as the grade received declines. In other words, among students who receive lower grades in BE, women choose to exit the field at an increasingly higher pace than their male counterparts. This likewise results in the increasing underrepresentation of women in BE as one moves down the grading scale.

The statistical significance of this gender imbalance in BE persistence is also depicted in Figure 1, represented by the rectangles. As with STEM, although not every point is statistically different from the next, the overall trend for persistence in BE does appear to be significant – that *male persistence dominates female persistence as the grade received in BE declines*. Furthermore, the empirical trend suggests that the degree to which male persistence dominates female persistence increases as the grade received declines.

When it comes to the trends in student persistence in SSH, a different situation emerges with regard to gender differences. The third column of the Social Science and Humanities-starting population panel of Table 7 shows that the statistically significant estimates in this column are all negative. This indicates that *women show relatively stronger persistence in SSH at those GPA levels.* This superiority of female persistence in SSH contrasts with the opposite STEM and BE patterns in gender differential in student persistence. This novel result shows that the phenomenon of relatively stronger sensitivity to grades exhibited by women in STEM and BE (which we reported above and is consistent with existing literature) is in fact discipline-specific, rather than being an unconditional characteristic of female students. A plausible interpretation of this new finding is that women, on average, may have relatively weaker attachments to STEM and BE disciplines, hence their stronger sensitivity to grades received in those fields.

The statistical significance of the gender imbalance in SSH persistence is illustrated in Figure 1, where it is represented by the squares. Although not each point is statistically different from the next, the overall trend for persistence in SSH does appear to be significant – that *dominance of female persistence over male persistence in SSH declines as the grade received in SSH declines, and eventually the gender of a student no longer has any statically significant impact on the persistence probabilities.*

To summarize the persistence results, the gender marginal effect estimates shown in Table 7 and illustrated by Figure 1 indicate that trends in student persistence and grade sensitivity require a more nuanced examination, disaggregated in terms of student academic concentrations. Indeed, although the trends uncovered for STEM and BE reveal female students to be increasingly responsive to declining grades in these categories in terms of inferior persistence relative to their male counterparts, this stronger sensitivity to grades on the part of women is *not* observed in SSH. There, women display relatively higher levels of persistence for the upper grade levels, but as the grades received decline, this relative dominance in persistence on the part of women fades and, eventually, at low grade levels, student's gender no longer impacts the probability of persisting in the field. This new finding suggests that stronger sensitivity to grades, rather than being a universal gender-specific phenomenon, is more likely to reflect gender differences in the underlying preferences for academic fields.

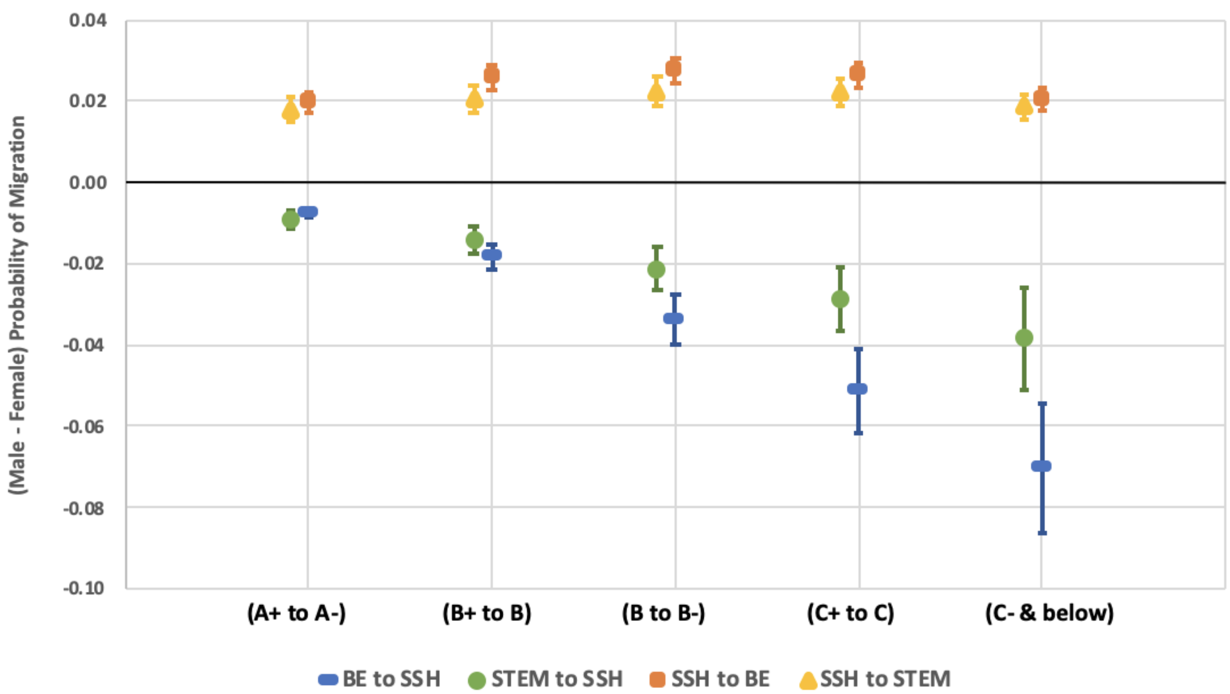
We now move to examine migration trends across academic categories.

Looking back at the transition probabilities in Table 6, a not-so-surprising, general trend emerges: as student grade level in a starting category decreases, the probability of migration to alternative academic categories increases, along with the probability of choosing to drop out. The exception to this trend is migration from STEM to BE, for which the probability of migration is fairly steady across all STEM grades received.

Looking across the STEM, BE and SSH starting groups there is no single 'most popular' migration destination. We also note that the popularity of a destination option fluctuates as the grade received declines, and these changes can differ between genders. For example, looking at the STEM-starting (the top) panel of Table 6, as the grade received in STEM declines, the probability of BE as a migration destination decreases while those for SSH or OP increases.

Looking across all three starting categories seen in Table 7, there are two features that remain consistent across the three academic categories we report on there. First, *men are consistently more likely to migrate to STEM and BE than women*, as indicated by the all-positive estimates seen in the SSH-starters' and BE-starters' STEM destination columns. Second, *women are consistently more likely to migrate to SSH than men*, as indicated by the all-negative estimates seen in the STEM-starters' and BE-starters' SSH destination columns. For this later migration trend, it is also worth noting that the gender differential grows as the grade received declines. In other words, women dominate the SSH migration pathway to a greater and greater extent as the grade received in the original category declines. Figure 2 illustrates the gender differences in along the migration pathways between SSH, STEM and BE.

Figure 2: Gender Differences in Migration Probabilities



The rectangles and circles seen in Figure 2 highlight the increasing popularity of SSH as a destination category for female students over male students. Note that the increasingly negative nature of the estimates signifies that the gender imbalance in each SSH migration group grows as the grade received in STEM or BE declines. For example, for BE-starters who earn a 'C- & below', the probability of a female student in this group choosing to switch from BE into SSH is nearly 7 percentage points higher than that of their male counterparts. Comparing this 7-percentage point difference to the less-than-1-percentage point difference associated with the 'A+ to A-' group of BE-starters, a clear widening of the gender differential emerges. A similar, but less dramatic widening of the gender differential is depicted for the STEM to SSH migration path shown in green. This empirical evidence demonstrates that SSH attracts women in relatively greater numbers, and that this effect strengthens as the grade received in STEM or BE declines. Thus, women not only are seen to show greater or equal persistence relative to men in SSH, but are also more likely to migrate to SSH than their male counterparts.

The migration path from SSH to BE, as well as the path from SSH to STEM do not showcase a widening gender differential (depicted as squares and triangles, respectively), but rather display a relatively constant one, such that being male increases one's likelihood of migration to either BE or STEM by approximately 2 percentage points. This empirical evidence demonstrates that STEM and BE attract men in relatively greater numbers, and that this effect is constant across all SSH grades received, controlling for all of our other regressors. This pattern clearly appears qualitatively consistent with the pattern of stronger persistence of men in STEM and/or BE as the starting field, as observed earlier. Thus, men not only are seen to show greater persistence than women in STEM and BE, but are also more likely to migrate to these fields from SSH than their female counterparts.

Together, the above comparisons of male and female persistence and migration trends across various academic pathways add up to an alternative interpretation of the relationship between students' sensitivity to grades and their attachment to a field of study, compared to that conjectured in the literature. Specifically, our results question the conjecture of stronger overall female sensitivity to grades as the factor driving them away from lower grades received in academic disciplines. Instead, our findings support a view that it is a student's underlying weaker preferential attachment to a field that makes him/her more sensitive to grades received in it, rather than a stronger sensitivity to grades on the part of a student (or a category of students) making them less persistent in fields which tend to assign lower grades.

Further, these results are consistent with conjectures that, for an array of societal reasons, there are persistent gender differences in the valuation of pecuniary and non-pecuniary characteristics of alternative careers. Indeed, such interpretations appear consistent with our findings of men's relative bias in the attachment to STEM and BE (which are, in fact, more lucrative), and the reversal or this trend in SSH, OP, and Education which are distinctly less lucrative and arguably associated with superior non-pecuniary benefits.

4.1 Robustness Checks

This subsection contains two robustness checks of our results presented in Section 4 on gender differences in persistence in and migration from academic disciplines. The first robustness check addresses the challenge of a potential role of peer effect in the results by introducing a *gender-peer effect* control variable into the regression. The second robustness check addresses the possibility that students' (unsuccessful) grade performance reflected in our category GPA regressor might be affected by their *ex ante* decision to *'give up'* on their current academic field before officially registering the change. Thus, the second robustness check – termed the *'Forfeiture'* check – runs the regression over the students' prior semester's grade information and examines the degree to which their estimated probabilities of persisting versus exiting changes. The new results produced from these robustness checks remain in line with our baseline results of Section 4. We shall now discuss the motivation, execution and outcomes of these robustness checks in more detail.

Gender-Peer Effects

A key baseline result of Section 4 shows that men are relatively more persistent than women in STEM and that this persistence differential grows as the grade received in STEM declines. It is plausible that the underlying underrepresentation of women in the field could itself contribute to their relatively lower persistence; in other words, the initial underrepresentation of one's gender peers might impact the student's decision whether to persist. In fact, based on such argument, the underrepresentation of women in STEM is often cited as a self-fulfilling phenomenon.

It is important to note that gender composition not only varies greatly across academic categories, but also amongst the majors within some of the academic categories under consideration. Gender variation among majors is particularly evident in STEM. For example, although women are underrepresented in the STEM category overall, the human biology BA/BS major is over 70% female, and the biology BA/BS major, one of the largest STEM major at Indiana University, is almost 52% female. However, computer science BA/BS and Physics BA/BS majors are each over 88% male.

Given the variation in gender composition within a given academic category, one can control for the potential *gender peer effects* on student's actions by averaging gender ratios over all classes taken by the student. This index captures the student's gender-peer experience in the category. In particular, we introduce a course section-specific *peer effects* variable. Since many courses are offered in multiple sections within a semester, this *gender peer effects* variable reports the gender composition specific to *each* section of relevant classes, as opposed to averaging across all sections of a class in a given semester. This approach is essential as it accounts for the relevant gender composition directly observed by affected students. Further, our *peer effects* variable is tied to the gender of a student the effect on whose actions is being estimated: if a student is male, then the peer effects variable will report the proportion of males in the class, *mutatis mutandis.* Once all the class section-specific gender ratios have been calculated for a student under consideration, we aggregate them within each academic category. We do so by grouping all the class sections the student has taken in accordance within each academic category and then take the average of these class section-specific gender ratios to represent the student's overall gender-peer experience by academic categories. This process thus yields five, student-specific, and category-specific *gender-peer effect* control variables.

Note that some students may not have taken courses in some categories, so there is no information regarding the peer-gender ratio in that category. For these students, a 'zero' entry for peer-gender ratio can be misleading for the task at hand. To address this issue, we include a *gender lacking-information dummy* for each academic category. This information dummy variable allows us to control the impact of the 'zero' gender-ratio entries resulting from a student under consideration not having taken any classes in the academic category on a student's category decision.

Table 8 presents the results of transition probability after controlling for these new *peer effect*, and *gender lacking-information dummy variables*. These results, similar to the baseline results, show the probability of persisting or exiting a given starting academic category for female and male students across a range of GPA levels.

Table 8: **Peer Effects: Female & Male Cross-Category Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	93.2%	95.4%	1.2%	1.6%	2.4%	1.6%	2.7%	1.0%	0.1%^{X}	0.0%^{X}	0.5%	0.4%
B+ to B	88.9%	92.9%	1.2%	1.7%	3.8%	2.6%	4.7%	1.9%	0.1%^{X}	0.0%^{X}	1.0%	0.9%
B to B-	83.1%	89.1%	1.0%	1.5%	6.1%	4.3%	7.1%	2.9%	0.3%*	0.1%^{X}	2.3%	2.2%
C+ to C	73.5%	81.5%	1.2%	1.8%	9.4%	6.9%	9.9%	4.2%	0.5%*	0.1%^{X}	5.5%	5.5%
C - and below	52.6%	61.5%	1.0%	1.5%	15.2%	12.0%	11.9%	5.4%	0.9%*	0.2%^{X}	18.4%	19.3%

Business & Economics-starting Population

BE Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	0.9%	1.4%	95.1%	96.2%	1.7%	0.9%	1.5%	0.9%	0.3%**	0.1%^{X}	0.6%	0.6%
B+ to B	1.7%	2.6%	88.3%	90.8%	3.9%	2.2%	4.2%	2.8%	0.2%**	0.0%^{X}	1.6%	1.6%
B to B-	2.9%	4.6%	76.6%	81.1%	7.1%	4.2%	9.6%	6.5%	0.4%**	0.1%*	3.4%	3.6%
C+ to C	4.4%	7.3%	59.4%	65.7%	12.7%	7.9%	15.3%	10.8%	0.8%*	0.2%*	7.4%	8.1%
C - and below	5.3%	9.2%	39.4%	45.5%	20.9%	14.1%	19.0%	14.0%	0.5%^{X}	0.1%^{X}	15.0%	17.0%

Social Sciences & Humanities-starting Population

SSH Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	2.7%	4.3%	1.4%	3.0%	89.3%	88.0%	4.6%	3.1%	0.7%	0.5%	1.3%	1.2%
B+ to B	2.9%	4.7%	1.9%	4.1%	83.0%	82.3%	8.6%	5.8%	0.9%	0.6%	2.8%	2.5%
B to B-	3.3%	5.4%	1.9%	4.4%	75.6%	75.6%	11.6%	8.0%	1.2%	0.7%	6.3%	5.8%
C+ to C	3.4%	5.6%	1.8%	4.2%	63.2%	64.2%	14.3%	10.1%	1.6%	1.0%	15.6%	14.7%
C - and below	2.4%	4.1%	1.3%	3.1%	40.8%	42.4%	12.7%	9.3%	2.1%	1.3%	40.8%	39.7%

X **not significant at 5% level, * 5% level, ** 1% level, all other results are significant at the 0.1% level.**

The probability estimates seen above in Table 8 are consistent with the baseline estimates displayed in Table 6. In fact, the Table 6 and 8 persistence estimates associated with higher grade levels (i.e., 'A+ to A-' and 'B+ to B') only differ by a fraction of a percentage point. Only when examining the estimates for the lower grade levels do we see a modest adjustment in values. For example, comparing the STEM-starting women from Table 6 and 8 in the 'C- & below' GPA range, we see that the two probability-of-persistence-in-STEM estimates differ by 1.1 percentage points (51.5% in Table 6 versus 52.6% in Table 8).

This latter evidence suggests that the gender-peer effects may matter less at higher grade levels, but may be stronger for students at lower grade levels. One can infer that a shortage of peers of a student's gender can magnify a negative effect of poor grades on his/her likelihood to persist in an academic category.

These *peer effect*-controlled results also sustain the second aspect of our baseline results of Section 4, namely the increase in gender differential in STEM and BE persistence associated with the decline in students' grade performance in these starting academic categories, as seen in Table 9.

Table 9: **Peer Effects: Gender Differences in Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	2.16%*** (0.00532)	0.00417 (0.00369)	-0.81%*** (0.00244)	-1.68%*** (0.00283)	-0.04% (0.000385)	-0.000405 (0.000685)
B+ to B	3.93%*** (0.00718)	0.00465 (0.00380)	-1.24%*** (0.00375)	-2.98%*** (0.00459)	-0.00109 (0.000747)	-0.07% (0.00147)
B to B-	5.97%*** (0.00963)	0.00445 (0.00327)	-1.87%** (0.00582)	-4.25%*** (0.00633)	-0.20% (0.00131)	-0.000963 (0.00333)
C+ to C	8.08%*** (0.0138)	0.58% (0.00392)	-2.57%** (0.00875)	-5.72%*** (0.00904)	-0.37% (0.00258)	-0.01% (0.00731)
C- & below	8.95%*** (0.0182)	0.55% (0.00333)	-3.23%* (0.0132)	-6.52%*** (0.0115)	-0.68% (0.00495)	0.94% (0.0175)

Business & Economics-starting Population

BE Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	0.47%* (0.00208)	1.07%* (0.00443)	-0.75%** (0.00230)	-0.53%** (0.00176)	-0.24% (0.00148)	-0.01% (0.00135)
B+ to B	0.95%* (0.00373)	2.43%** (0.00918)	-1.71%** (0.00530)	-1.47%** (0.00488)	-0.22% (0.00143)	0.02% (0.00334)
B to B-	1.73%** (0.00594)	4.49%** (0.0158)	-2.93%** (0.00933)	-3.11%** (0.0106)	-0.35% (0.00211)	0.17% (0.00704)
C+ to C	2.92%*** (0.00877)	6.28%** (0.0216)	-4.73%** (0.0157)	-4.53%** (0.0163)	-0.64% (0.00411)	0.70% (0.0138)
C- & below	3.91%*** (0.0111)	6.13%** (0.0216)	-6.77%** (0.0232)	-4.92%* (0.0199)	-0.41% (0.00325)	2.07% (0.0236)

Social Sciences & Humanities-starting Population

SSH Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	1.58%*** (0.00330)	1.60%*** (0.00274)	-1.28%** (0.00483)	-1.51%*** (0.00244)	-0.26%* (0.00104)	-0.12% (0.00103)
B+ to B	1.78%*** (0.00346)	2.23%*** (0.00337)	-0.70% (0.00630)	-2.73%*** (0.00435)	-0.34%* (0.00138)	-0.24% (0.00211)
B to B-	2.09%*** (0.00380)	2.41%*** (0.00338)	-0.01% (0.00790)	-3.57%*** (0.00579)	-0.42%* (0.00180)	-0.49% (0.00464)
C+ to C	2.24%*** (0.00393)	2.40%*** (0.00338)	1.01% (0.0107)	-4.15%*** (0.00726)	-0.60%* (0.00274)	-0.89% (0.00989)
C- & below	1.70%*** (0.00315)	1.83%*** (0.00293)	1.66% (0.0127)	-3.43%*** (0.00735)	-0.73% (0.00384)	-1.02% (0.0161)

Standard errors in parentheses
***** $p < 0.05$, ****** $p < 0.01$, ******* $p < 0.001$

When examining the columns of Table 9 characterizing gender differentials in persistence in STEM and BE, we see all positive values that increase in magnitude as the grade received declines. These STEM and BE persistence columns indicate that male students dominate women, in terms of persistence probability, in each GPA bin, and that furthermore, the differential increases as the GPA declines. These trends of growing male prevalence in persistence in STEM and BE preserve our finding in the baseline results of Section 4. Although one can note a minor moderation in the strength of this effect of declining GPA in Table 9, it is clearly sustained under this robustness check.

To conclude the analysis of the gender-peer robustness check, we observe that the results in Table 9 remain consistent with the baseline results of Section 4 also in our analysis of student persistence in the SSH category. Namely, the new, peer effect controlled results affirm the *absence* of male-dominated differential in persistence in SSH. Just as Table 7 of Section 4, Table 9 shows that being male does not increase a student's probability of persisting in SSH to any statistically significant degree. In fact, in Table 9 one can see that *being female increases the probability of persisting in SSH* by 1.28 percentage points when in the 'A+ to A-' GPA-range. Then, as the grade-level declines, *student gender no longer shows any statistically significant impact on the probability of persisting in SSH.*

'Forfeiture' Robustness Check

As stated previously, the second robustness check addresses the possibility that students might set their minds on *'giving up'* on their current major/academic field before they officially change their major. In such instances, it is possible that students who enrolled in some classes associated with the original major before having formed plans to change the major, will not immediately register the change in the course of the semester while still needing to complete the classes in which they are already enrolled. In such cases, students' weak performance in such classes may constitute 'forfeiture', i.e., stem from their decision to abandon the corresponding major, rather than serving as its cause.

To address this potential issue, we perform our second, *'forfeiture'* robustness check, we look one semester back, before the student's final semester in a given category, run the regression over the students' penultimate semester's grade information, and produce new estimates of the probabilities of persistence and migration. Below, we will compare these new, *'forfeiture'* regression estimates with the baselines estimates.

Table 10, otherwise similar to Table 6, displays the estimated probabilities of female and male students persisting in or transitioning from a given academic category to an alternative using this penultimate GPA information.

Table 10: **Forfeiture: Female & Male Cross-Category Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	93.2%	95.3%	1.0%	1.5%	2.6%	1.6%	2.5%	1.0%	0.0%^{X}	0.0%^{X}	0.6%	0.6%
B+ to B	89.2%	93.0%	1.0%	1.4%	4.0%	2.5%	4.3%	1.8%	0.1%^{X}	0.0%^{X}	1.3%	1.3%
B to B-	83.3%	88.9%	1.0%	1.5%	6.2%	4.0%	6.5%	2.8%	0.2%^{X}	0.1%^{X}	2.8%	2.7%
C+ to C	74.1%	81.8%	1.1%	1.7%	9.1%	6.1%	9.2%	4.1%	0.4%^{X}	0.1%^{X}	6.2%	6.2%
C - and below	54.8%	64.1%	1.3%	2.0%	15.8%	11.6%	12.3%	5.9%	0.8%^{X}	0.3%^{X}	14.9%	16.1%

Business & Economics-starting Population

BE Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	1.1%	1.3%	94.5%	96.0%	1.8%	0.9%	1.7%	1.0%	0.2%*	0.0%^{X}	0.8%	0.8%
B+ to B	2.1%	2.5%	87.5%	90.6%	4.1%	2.2%	4.4%	2.7%	0.2%**	0.0%*	1.8%	1.9%
B to B-	3.5%	4.5%	75.8%	81.3%	7.8%	4.4%	9.1%	6.0%	0.3%**	0.1%*	3.5%	3.8%
C+ to C	4.9%	6.7%	59.5%	67.2%	13.4%	8.0%	15.0%	10.4%	0.6%*	0.1%*	6.6%	7.5%
C - and below	6.0%	8.7%	39.7%	47.7%	21.3%	13.9%	20.8%	15.5%	0.6%*	0.2%^{X}	11.6%	14.0%

Social Sciences & Humanities-starting Population

SSH Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	2.4%	4.2%	1.2%	2.8%	90.0%	87.9%	4.2%	3.2%	0.7%	0.3%	1.5%	1.6%
B+ to B	2.7%	4.7%	1.7%	3.9%	84.1%	82.0%	7.6%	5.8%	1.0%	0.4%	3.1%	3.1%
B to B-	2.8%	5.0%	1.9%	4.4%	77.4%	75.6%	10.4%	7.9%	1.2%	0.5%	6.4%	6.6%
C+ to C	2.7%	4.8%	1.9%	4.5%	67.3%	66.2%	12.6%	9.7%	1.9%	0.9%	13.5%	14.0%
C - and below	2.2%	4.0%	1.7%	4.1%	49.2%	48.6%	14.3%	11.0%	2.3%	1.1%	30.2%	31.2%

X **not significant at 5% level, * 5% level, ** 1% level, all other results are significant at the 0.1% level.**

The probability estimates seen in Table 10 are consistent with the baseline estimates displayed in Table 6. While the former feature, at times, modest differences from the latter, the persistence and migration patterns for female and male students are the same. In particular, persistence in each category declines as the grade received decreases. SSH and OP remain popular migration destination categories for both male and female students, particularly at the lower grade levels. In order to trace explicit comparisons between gender differentials in persistence and migration decisions we examine the gender marginal effects. The results are presented in Table 11.

Table 11, otherwise similar to Table 7, shows the impact of a student's gender on the probability of persisting or migrating to an academic category, while performing the *'forfeiture'* check. The positive values continue to signify that being male increases the likelihood of a given action.

Table 11: **Forfeiture: Gender Differences in Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	2.16%***	0.43%	-1.01%***	-1.49%***	-0.03%	-0.05%
	(0.00448)	(0.00265)	(0.00241)	(0.00245)	(0.000288)	(0.000828)
B+ to B	3.74%***	0.43%	-1.48%***	-2.53%***	-0.08%	-0.08%
	(0.00606)	(0.00252)	(0.00349)	(0.00375)	(0.000595)	(0.00180)
B to B-	5.66%***	0.48%	-2.18%***	-3.72%***	-0.16%	-0.08%
	(0.00862)	(0.00258)	(0.00530)	(0.00538)	(0.00102)	(0.00361)
C+ to C	7.68%***	0.57%*	-2.92%***	-5.09%***	-0.28%	0.04%
	(0.0124)	(0.00284)	(0.00759)	(0.00780)	(0.00188)	(0.00764)
C- & below	9.26%***	0.78%*	-4.21%***	-6.45%***	-0.53%	1.15%
	(0.0166)	(0.00347)	(0.0128)	(0.0113)	(0.00403)	(0.0154)

Business & Economics-starting Population

BE Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	0.214%	1.45%***	-0.85%***	-0.68%***	-0.14%	-0.002%
	(0.00152)	(0.00302)	(0.00153)	(0.00144)	(0.000719) .	(0.00113)
B+ to B	0.49%	3.16%***	-1.89%***	-1.65%***	-0.15%*	0.05%
	(0.00289)	(0.00612)	(0.00340)	(0.00355)	(0.000746)	(0.00252)
B to B-	1.03%*	5.50%***	-3.43%***	-3.15%***	-0.21%	0.26%
	(0.00477)	(0.0101)	(0.00640)	(0.00708)	(0.00112)	(0.00473)
C+ to C	1.80%**	7.65%***	-5.32%***	-4.60%***	-0.43%	0.90%
	(0.00671)	(0.0135)	(0.0107)	(0.0113)	(0.00258)	(0.00853)
C- & below	2.70%**	7.96%***	-7.33%***	-5.35%***	-0.45%	2.47%
	(0.00847)	(0.0138)	(0.0163)	(0.0156)	(0.00347)	(0.0139)

Social Sciences & Humanities-starting Population

SSH Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	1.86%***	1.61%***	-2.10%***	-1.02%***	-0.40%***	0.05%
	(0.00314)	(0.00231)	(0.00463)	(0.00231)	(0.000996)	(0.00123)
B+ to B	2.08%***	2.23%***	-2.04%***	-1.85%***	-0.52%***	0.09%
	(0.00331)	(0.00286)	(0.00613)	(0.00407)	(0.00126)	(0.00238)
B to B-	2.21%***	2.50%***	-1.76%*	-2.50%***	-0.63%***	0.19%
	(0.00338)	(0.00303)	(0.00780)	(0.00553)	(0.00161)	(0.00473)
C+ to C	2.13%***	2.60%***	-1.19%	-2.96%***	-1.02%***	0.45%
	(0.00326)	(0.00317)	(0.0102)	(0.00677)	(0.00295)	(0.00904)
C- & below	1.78%***	2.35%***	-0.63%	-3.29%***	-1.26%**	1.04%
	(0.00305)	(0.00337)	(0.0127)	(0.00813)	(0.00441)	(0.0150)

Standard errors in parentheses
***** $p < 0.05$, ****** $p < 0.01$, ******* $p < 0.001$

The all-positive values in the columns associated with STEM and BE throughout Table 11 indicate that male students remain more likely to persist in or migrate to STEM and BE than their female counterparts, and that this is true across all GPA ranges. Also, this gender differential in persistence in favor of men in these categories continues to grow as the grade received declines. These trends affirm those found in the baseline results presented in Table 7 of Section 4.

The negative values throughout the SSH and OP columns of Table 11 indicate that female students are not only more likely to persist in SSH than their male counterparts (at certain grade levels), but also more likely to migrate to SSH, OP, and EDUC. The one caveat for the former part of this statement is that at lower grade levels in the SSH category, being female no longer increases student's likelihood to persisting in it. In fact, a student's gender no longer has any statistically significant impact on the probability of persisting in SSH once we enter the 'C+ to C' and below GPA-ranges. These same trends are also found in the baseline results seen in Table 7, Section 4.

To summarize, the patterns presented in Table 11 concerning gender differentials in persistence in or migration across academic categories are in perfect agreement with the baselines results of Section 4.

Thus, overall, while the *'forfeiture'* robustness check estimates presented in Tables 10 and 11 feature modest differences from the baseline results of Section 4, the trends and patterns therein are again sustained when the robustness control is applied. So is the broader conclusion of our analysis that stronger sensitivity to grades, rather than being a gender-specific phenomenon, is more likely to reflect gender differences in the underlying preferences for academic fields.

5 Trade-offs in Choosing a College Major: a Theoretical Framework

One can think of an individual college student's decision to choose an academic field and then either to persist in it or to switch to a different one in the spirit of Manski (1989) who analyzed sequential individual decisions about whether to enroll and then to persist in college. In such a utilitarian framework, students consider characteristics of a major, as well as their own individual characteristics, such as abilities and tastes. Recent research shows that expected future career earnings prominently enter students' preferences (see Arcidiacono, 2004), as does the effort required to pursue the corresponding studies (e.g., Montmarquette et al. and Babcock and Marks, *op cit.*).

Thus, each student i can be seen as making educational choices to maximize utility function $U_i(I, e, g, \Theta|a(i))$. The student's utility increases in pecuniary benefit I (i.e., the expected career income) of the corresponding career path, in its non-pecuniary benefits Θ, and in the GPA grade g earned, and decreases in the required effort e, all conditional on student characteristics $a(i)$ such as ability. Assuming separability, for the simplicity of argument but without loss of generality of the analytical framework, one can write:

$$max U_i\big(I, e, \Theta|a(i)\big) = \alpha_i u(I) - \beta_i v\big(e|a(i)\big) + \gamma_i s(\Theta) + t(g), \tag{1}$$

where the first two components are associated with income and effort, the third one represents non-pecuniary benefits associated with an educational choice (including the attributes of an academic subject and the corresponding career), and the last the component accounts for the psychic benefits derived from receiving a high grade and, accordingly, the distress from ("sensitivity" to) a low grade. All of the components are conditional on student's ability, since student's future career success, effort required to meet educational requirements, and even non-pecuniary benefits of a major and future career all depend on it. The coefficients α_i, β_i, and γ_i stand for the weights student i attaches to the components of welfare associated with income, effort, and non-pecuniary benefits of a major choice, respectively, relative to that associated with the GPA grade received.

To impose further structure, consistent with realities of choices in higher education, let us posit that career income derived by student i from a degree in major $m = 1, 2, \ldots, M$ is given by $w_m h_m(i)$ where w_m is the exogenously given wage rate per unit of human capital and $h_m(i)$ is the attainment level of the major-specific human capital chosen by the student.

We further posit that each major establishes a ladder of its academic standards corresponding to each grade, such that $h_{m,g}$ is the level of attainment of human capital specific to major m required for obtaining grade g in the major. The human capital attainment choices made by a student pursuing major m are then reduced to the choice of grade he/she wishes to receive and, by implication, the labor market benefits this will provide and the effort this will require from this individual student. Further reducing, for the sake of simplicity, individual characteristics to a single parameter of universal ability a and assuming a specific functional form relating the effect of ability on educational effort (again, with no loss for the generality of the argument), we can reduce the maximization problem (1) to the following:

$$\begin{aligned} max_{m,g} U_i(m,g) &= U_i\Big(I(m,g), e(m,g|a(i)), \Theta(m), g)\Big) \\ &= \alpha_i u(w_m h_{m,g}) - \beta_i v\left(\frac{h_{m,g}}{a(i)}\right) + \gamma_i s(\Theta(m)) + t(g) \end{aligned} \tag{2}$$

which indeed implies that each student's choices can be reduced to the choices of major m, and the grade level g to pursue in it.

It is important to note that if students do not have advance knowledge of their ability levels, the above optimization should be stated in expected utility form, proceeding from a posterior distribution of ability based on a signal obtained prior to decision-making. Once an updated signal is obtained, a student may reassess the optimization accordingly, which may result in changes in educational choices, such as those studied in this paper.

Given the exogenous wage rates, one can call major m more lucrative than major j if $w_m > w_j$. It then follows, from the empirical evidence and model-based reasoning presented by Kaganovich and Su (2018)[13], that for each grade level g the corresponding academic standards compare in the same order in equilibrium: $h_{m,g} > h_{j,g}$, so attaining a particular grade level in a more lucrative major will take more effort than in a less lucrative one. We conjecture that these results extend to this broader framework with many majors and grade levels and formulate this as the following condition:

<u>Condition 1:</u> The more lucrative a major, the stricter are its grading standards. Namely, let major m be more lucrative than major j, i.e., $w_m > w_j$. Then for each grade level g the corresponding academic standards compare as follows: $h_{m,g} > h_{j,g}$, such that attaining the grade takes more effort in the more lucrative major m.

Assuming that problem (2) satisfies requisite sufficient conditions of optimality, the first order conditions then imply the following lemma:

<u>Lemma 1:</u> For each pair of majors m, j, such that $w_m > w_j$, there is a cut-off ability level

$$a(m,j) = a(m,j|\alpha_i, \beta_i, \gamma_i) \tag{3}$$

such that, for a given set of weight coefficients $\alpha_i, \beta_i, \gamma_i$, students with ability below $a(m,j)$ will prefer the less lucrative of the two majors, j, while students with ability above the threshold will prefer major m in this binary choice.

[13]Kaganovich and Su (2018) present empirical evidence of a trade-off between career earnings associated with college majors and their academic standards (see also Diette and Raghav, 2016): namely, more lucrative majors tend to maintain less generous grading standards; conversely, less lucrative majors are more prone to grade inflation. They then develop an explanation of this phenomenon through a model of intra-university competition between departments for students, consistent with some ideas proposed by Aachen and Courant (2009). The academic standards for grades are chosen by individual departments and serve as their competitive instruments for attracting students to the corresponding disciplines or for deterring them, depending on academic eligibility. Since grading standards determine students' effort for attaining them according to individual levels of prior preparation (ability), students then decide which levels of effort are worth the monetary rewards associated with the alternative educational choices, hence the trade-offs between the standards and the rewards across the disciplines. This leads to student self-sorting across the departments, and, in intra-university competitive equilibrium, to a positive relationship between job market rewards of majors and their grading standards, as described above. Indeed, in their competition for students, departments whose majors offer inferior job market rewards must choose lower academic standards, i.e., "inflate" their grades, to allow their students less required effort as a compensation.

It is then straightforward to derive from the above Lemma a further crucial insight:

Lemma 2: The ability threshold $a(m, j)$ for choosing more lucrative major m over the less lucrative j declines in the weight the student places on pecuniary benefit of a major, α_i, and increases in the weight he/she places on the effort associated with the chosen human capital attainment level, β_i.

This last result allows one to conclude, in this stylized analytical framework, that students who value pecuniary benefit of a degree more, will not only prefer a higher paying major *ceteris paribus*, but that *ability sorting among students who value pecuniary benefits more strongly, will be more strongly biased in favor of ("attached" to) a lucrative major.* A further implication of the above finding is that there is a *trade-off between lucrativity of a major and a grade level acceptable to student*; in fact, the higher the weight a student of given ability attaches to pecuniary benefits of a major, the lower the acceptable grade level, at which this student will still prefer the choice of the more lucrative major. In other words, students who place a higher weight on future income (which is more common for men, according to the above referenced studies) will exhibit relatively less sensitivity to grades in majors that are more lucrative than their peers (predominantly women) who are less attracted to pecuniary benefits. In contrast, the former type students (i.e., most men) will exhibit stronger sensitivity to grades when pursuing a less lucrative major.

6 Concluding Comments

Our results affirm that male students exhibit significantly stronger persistence than women, *ceteris paribus*, in STEM and BE disciplines, and that this gender differential is the strongest among students receiving low grades in STEM and/or the Business and Economics (BE) category. Among students starting, but not persisting in STEM or BE, women dominate in migration to Social Sciences and Humanities (SSH) in particular. This dominance, again, strengthens as grades in the starting category decline. These results appear consistent with what the literature has characterized as stronger "sensitivity" to grades among women.

Importantly, however, we were able to demonstrate that this pattern of superior persistence on the part of men does not extend to students whose starting category is SSH. We have shown that, in this academic category, women exhibit greater or equal persistence relative to their male counterparts. In fact, at the lowest grade levels in SSH, student's gender no longer has a statistically significant impact on the probability of persistence in SSH. This somewhat unexpected result is a distinctly novel contribution to the literature on the subject. We have been able to obtain it owing to the uniquely rich IU Learning Analytics data, which allowed us to look at student transitions across disciplines, whereas existing literature primarily focuses on persistence in STEM category alone.

These results support our argument that stronger sensitivity to grades, rather than being a gender-specific phenomenon, is more likely to reflect gender differences in the underlying preferences for academic fields, whose existence in principle has been documented in the literature. Furthermore, our results suggest that it is the lower underlying preference of a student for a field of study that is likely to make him or her more "sensitive" to grades received in it, rather than the other way around. In other words, this runs counter to a commonly suggested understanding that it is innate stronger sensitivity to grades characteristic of student types that makes them less attached to academic disciplines, which are known to assign lower grades.

While studying the factors behind student persistence in or attrition from starting academic categories, we focused on student migration to alternative categories without a detailed analysis of the factors determining students' decisions to drop out of IU altogether. As explained in Section 2, an examination of drop out decisions is best done within the first year of a students' academic career. Indeed, the overwhelming majority of dropout decisions occurs then, as opposed to at the end of the second year where we focus our analysis in this paper, which, as we explained, is the prevalent time when students formalize their selection of an academic discipline. Our present study, however, offers insights for fruitful novel lines of inquiry into students' dropout decisions. As well articulated by Manski (1989), a decision to drop out reflects a student's assessment that his/her expected welfare value of the outside options is superior to that associated with persisting in college given the student's performance. This then implies that student's valuation of alternative disciplines offered by college, along with his/her current discipline must have bearing for a dropout decision, controlling for performance. For instance, a student pursuing STEM or BE and receiving poor grades there, is more likely to decide to drop out rather than switch to “softer” alternatives, if he/she attaches low welfare values to the latter. The insights developed in this paper allow us to conjecture, in particular, that men who perform poorly in STEM or BE, enough to decide to exit these disciplines, are more likely to drop out of the university rather than switch to an academic alternative such as SSH, relative to comparable women, controlling for other relevant characteristics. Exploring this novel conjecture is a subject matter of a separate study, which IU Learning Analytics data affords us to pursue.

References

[1] Aachen, A.C. and P.N. Courant (2009). "What Are Grades Made of," *Journal of Economic Perspectives* 23, 77-92.

[2] Altonji, J.G., P. Arcidiacono, and A. Maurel (2015) "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects," NBER Working Paper 21655.

[3] Arcidiacono, P. (2004). "Ability sorting and the returns to college major," *Journal of Econometrics* 121, 343-375.

[4] Arcidiacono P., V.J. Hotz, and S. Kang (2012). "Modeling college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics* 166, 3–16.

[5] Ashworth J. and T. Ransom (2019). "Has the college wage premium continued to rise? Evidence from multiple U.S. surveys," *Economics of Education Review* 69, 149-154.

[6] Babcock, P. and M. Marks (2011). "The Falling Time Cost of College: Evidence from Half a Century of Time Use Data," *Review of Economics and Statistics* 93, 468-478.

[7] Becker, G.S., W.H.J. Hubbard, and K.M. Murphy (2010). "Explaining the Worldwide Boom in Higher Education of Women," *Journal of Human Capital* 4, 203-241.

[8] Berger, M.C. (1988). "Predicted Future Earnings and the Choice of College Major," *Industrial and Labor Relations Review* 41, 418-429.

[9] Chen, X. and M. Soldner (2013). "STEM Attrition: College Students' Path Into and Out of STEM Fields. Statistical Analysis Report." NCES 2014-001, National Center for Education Statistics, Institute of Education Sciences. U.S. Department of Education.

[10] Feng, Y., S. Fiorini, L. Shepard, and D. Groth (2018). "Retention of Students in STEM Disciplines: A Closer Look at Gender and Academic Performance." Indiana University mimeo.

[11] Gemici, A. and M. Wiswall (2014). "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors," *International Economic Review* 55, 23–56.

[12] Goldin, C., L. Katz, and I. Kuziemko (2006). "The Homecoming of American College Women: the Reversal of the College Gender Gap," *Journal of Economic Perspectives* 20, 133–156.

[13] Huang, G., N. Taddese, and E. Walter (2000). "Entry and Persistence of Women and Minorities in College Science and Engineering Education." Washington, DC. National Center for Education Statistics, NCES 2000601.

[14] Hershbein, B. and M.S. Kearney (2014). "Major Decisions: What Graduates Earn over Their Lifetimes." The Hamilton Project Papers. The Brookings Institution. `http://www.hamiltonproject.org/papers/major_decisions_what_graduates_earn_over_their_lifetimes`

[15] James, J. (2012). "The College Wage Premium," Economic Commentary, 2012-10, The Federal Reserve bank of Cleveland.

[16] Kaganovich, M. and X. Su (2018). "Grade-Compensating Differentials in the Competition between College Majors." Indiana University mimeo.

[17] Kirkeboen, L.J., E. Leuven, and M. Mogstad (2016) "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics* 131, 1057-1111.

[18] Manski, C. F., (1989). "Schooling as Experimentation: a Reappraisal of the Postsecondary Dropout Phenomenon," *Economics of Education Review* 8, 305-312.

[19] Montmarquette, C., K. Cannings, and S. Mahseredjian (2002) "How do young people choose college majors?" *Economics of Education Review* 21, 543-556

[20] Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29(6), 923-934.

[21] Rask, K. (2010). "Attrition in STEM fields in a liberal arts college: The importance of grades and pre-collegiate preferences," *Economics of Education Review* 29, 892-900.

[22] Rask, K. and J. Tiefenthaler (2008). "The Role of Grade Sensitivity in Explaining the Gender Imbalance in Undergraduate Economics," *Economics of Education Review* 27, 676-687.

[23] Rendall, A., and M. Rendall (2014). "Math Matters: Education Choices and Wage Inequality." University of Zurich working paper No. 160.

[24] Stinebrickner, R. and T.R. Stinebrickner (2014). "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout," *Review of Economic Studies* 81, 426-472.

[25] Zafar, B. (2013). "College Major and the Gender Gap," Journal of Human Resources 48, 545-595.

Appendix

Chart 1: Departments Included in each Academic Category

STEM: see Homeland Security STEM field classifications

Business and Economics: BL-BUS, BL-ECON

Education: BL-EDUC

Social Sciences and Humanities:

BL-AFRI	BL-COGS	BL-INTL	BL-JOUR	BL-POLS
BL-AFRO	BL-EALC	BL-ISLM	BL-JSP	BL-PSY
BL-AMST	BL-ENG	BL-FOLK	BL-LATS	BL-REEI
BL-ANTH	BL-EURO	BL-FRIT	BL-LBST	BL-REL
BL-ARSC	BL-FOLK	BL-GEOG	BL-LING	BL-RENA
BL-ARTH	BL-FRIT	BL-GERM	BL-LTAM	BL-SEAS
BL-CEUS	BL-GEOG	BL-GNDR	BL-MEST	BL-SEMS
BL-CJUS	BL-GERM	BL-HIST	BL-MSCH	BL-SLAV
BL-CLAR	BL-GNDR	BL-HPSC	BL-MUS	BL-SOC
BL-CLAS	BL-HIST	BL-INDI	BL-MUSDPT	BL-SPAN
BL-CMCL	BL-HPSC	BL-INTL	BL-NELC	BL-TELC
BL-CMLT	BL-INDI	BL-ISLM	BL-PHIL	BL-THTR
				BL-VICT

Other Professional:

BL-AADM	BL-SPEA
BL-SOAD	BL-SPH
BL-FINA	BL-SWK
BL-AMID	BL-DENT
BL-HPER	BL-MED
BL-INFO	BL-NURS
BL-SPCN	BL-OPT

Chart: Timing of Dropout

Timing of Drop	Number	% of total
1st year	5,586	54.0%
2nd year	2,543	24.6%
3rd year	1,008	9.7%
4th year	758	7.3%
5th year	349	3.4%
6th year	69	0.7%
7th year	40	0.4%
Total drops	**10,353**	**100.0%**

Table A: **Female & Male Cross-Category Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	93.0%	95.4%	1.3%	1.5%	2.5%	1.6%	2.7%	1.1%	0.0%^{X}	0.0%^{X}	0.5%	0.4%
B+ to B	88.9%	93.2%	1.3%	1.5%	3.9%	2.5%	4.7%	1.9%	0.1%^{X}	0.0%^{X}	1.0%	0.9%
B to B-	82.8%	89.3%	1.2%	1.5%	6.3%	4.1%	7.1%	2.9%	0.2%*	0.1%^{X}	2.5%	2.1%
C+ to C	73.2%	82.1%	1.3%	1.6%	9.3%	6.4%	9.7%	4.2%	0.5%**	0.1%^{X}	6.1%	5.6%
C - and below	51.5%	61.9%	1.1%	1.5%	15.9%	12.1%	11.8%	5.5%	0.8%*	0.3%^{X}	18.9%	18.8%

Business & Economics-starting Population

BE Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	1.0%	1.2%	95.0%	96.5%	1.5%	0.8%	1.7%	1.0%	0.2%*	0.0%^{X}	0.6%	0.6%
B+ to B	2.1%	2.6%	87.5%	90.9%	4.0%	2.1%	4.5%	2.8%	0.2%**	0.0%*	1.6%	1.5%
B to B-	3.5%	4.4%	75.3%	81.4%	7.8%	4.4%	9.4%	6.1%	0.4%**	0.1%*	3.6%	3.6%
C+ to C	5.0%	6.8%	58.4%	66.7%	13.0%	7.9%	15.3%	10.4%	0.6%*	0.1%*	7.8%	8.1%
C - and below	6.2%	9.0%	37.4%	45.5%	20.7%	13.7%	20.1%	14.8%	0.4%^{X}	0.1%^{X}	15.2%	16.9%

Social Sciences & Humanities-starting Population

SSH Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	2.6%	4.4%	1.3%	3.3%	89.5%	87.4%	4.5%	3.4%	0.7%	0.4%	1.3%	1.2%
B+ to B	3.0%	5.0%	1.7%	4.3%	83.3%	81.5%	8.2%	6.1%	1.0%	0.5%	2.8%	2.5%
B to B-	3.2%	5.4%	1.8%	4.6%	76.3%	75.2%	11.0%	8.3%	1.3%	0.6%	6.4%	5.9%
C+ to C	3.0%	5.3%	1.7%	4.3%	64.2%	64.4%	13.5%	10.4%	2.0%	1.0%	15.6%	14.6%
C - and below	2.4%	4.3%	1.2%	3.3%	40.6%	41.6%	12.7%	10.0%	2.2%	1.1%	40.9%	39.6%

Other Professional-starting Population

OP Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	1.6%	2.1%	0.8%	2.8%	2.9%	3.1%	92.0%	90.0%	0.4%	0.2%**	2.5%	1.8%
B+ to B	1.3%	1.8%	1.0%	3.7%	4.0%	4.4%	87.1%	85.5%	0.7%	0.3%**	5.9%	4.3%
B to B-	1.3%	1.8%	1.1%	4.0%	5.7%	6.4%	78.0%	77.9%	1.2%	0.6%**	12.7%	9.4%
C+ to C	1.3%	1.9%	1.0%	3.7%	7.0%	8.0%	65.3%	67.1%	1.8%	0.9%**	23.6%	18.5%
C - and below	1.0%**	1.5%**	0.7%**	3.0%	7.2%	8.8%	41.3%	45.2%	2.4%*	1.2%*	47.4%	40.3%

Education-starting Population

EDUC Grade	STEM		BE		SSH		OP		Educ		DROP	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
A+ to A-	0.4%*	2.1%*	0.1%^{X}	1.1%*	3.0%	6.1%	3.5%	3.3%	91.4%	86.1%	1.6%	1.3%**
B+ to B	0.4%*	2.2%*	0.2%^{X}	1.6%^{X}	4.8%	9.5%	6.1%	5.7%	82.1%	76.0%	6.4%	5.0%
B to B-	0.4%^{X}	1.8%^{X}	0.5%^{X}	3.4%^{X}	6.6%	12.7%	7.7%	7.1%**	66.2%	60.4%	18.6%	14.5%
C+ to C	0.2%^{X}	1.3%^{X}	0.2%^{X}	2.6%^{X}	6.7%**	13.2%	5.9%**	5.8%**	52.2%	48.6%	34.8%	28.4%
C - and below	0.1%^{X}	0.5%^{X}	0.3%^{X}	2.6%^{X}	6.1%*	12.4%**	3.0%^{X}	2.9%^{X}	22.7%	21.2%	67.8%	60.4%

X not significant at 5% level, * 5% level, ** 1% level, all other results are significant at the 0.1% level.

Table B: **Gender Differences in Transition and Persistence Probabilities**

STEM-starting Population

STEM Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	2.48%*** (0.00461)	0.23% (0.00292)	-0.936%*** (0.00229)	-1.65%*** (0.00260)	-0.04% (0.000281)	-0.08% (0.000611)
B+ to B	4.26%*** (0.00618)	0.25% (0.00287)	-1.40%*** (0.00340)	-2.85%*** (0.00402)	-0.09% (0.000625)	-0.17% (0.00136)
B to B-	6.49%*** (0.00870)	0.28% (0.00274)	-2.13%*** (0.00531)	-4.15%*** (0.00570)	-0.17% (0.000991)	-0.33% (0.00313)
C+ to C	8.89%*** (0.0125)	0.36% (0.00293)	-2.85%*** (0.00773)	-5.52%*** (0.00797)	-0.32% (0.00197)	-0.56% (0.00722)
C- & below	10.40%*** (0.0168)	0.41% (0.00266)	-3.85%** (0.0128)	-6.26%*** (0.0106)	-0.54% (0.00387)	-0.16% (0.0168)

Business & Economics-starting Population

BE Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	0.18% (0.00136)	1.43%*** (0.00269)	-0.74%*** (0.00134)	-0.68%*** (0.00135)	-0.14%* (0.000679)	-0.06% (0.000827)
B+ to B	0.47% (0.00285)	3.42%*** (0.00598)	-1.84%*** (0.00328)	-1.77%*** (0.00354)	-0.18%* (0.000810)	-0.10% (0.00218)
B to B-	0.99%* (0.00463)	6.11%*** (0.0102)	-3.38%*** (0.00631)	-3.36%*** (0.00706)	-0.31%* (0.00148)	-0.04% (0.00469)
C+ to C	1.78%** (0.00669)	8.30%*** (0.0137)	-5.13%*** (0.0103)	-4.86%*** (0.0112)	-0.45% (0.00270)	0.34% (0.00928)
C- & below	2.78%** (0.00851)	8.16%*** (0.0134)	-7.03%*** (0.0158)	-5.30%*** (0.0147)	-0.34% (0.00279)	1.73% (0.0157)

Social Sciences & Humanities-starting Population

SSH Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	1.78%*** (0.00315)	1.98%*** (0.00255)	-2.10%*** (0.00469)	-1.17%*** (0.00238)	-0.39%*** (0.000990)	-0.11% (0.000972)
B+ to B	2.06%*** (0.00342)	2.59%*** (0.00298)	-1.82%** (0.00621)	-2.08%*** (0.00420)	-0.52%*** (0.00128)	-0.24% (0.00204)
B to B-	2.23%*** (0.00353)	2.76%*** (0.00300)	-1.10% (0.00784)	-2.71%*** (0.00563)	-0.67%*** (0.00178)	-0.51% (0.00450)
C+ to C	2.23%*** (0.00341)	2.64%*** (0.00297)	0.16% (0.0106)	-3.09%*** (0.00697)	-1.00%*** (0.00300)	-0.94% (0.00956)
C- & below	1.86%*** (0.00306)	2.05%*** (0.00289)	1.09% (0.0123)	-2.65%*** (0.00726)	-1.10%** (0.00397)	-1.25% (0.0157)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B Continued: **Gender Differences in Transition and Persistence Probabilities**

Other Professional-starting Population

OP Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	0.00590 (0.00355)	0.0200*** (0.00338)	0.00248 (0.00360)	-0.0191** (0.00637)	-0.00228* (0.000902)	-0.00697*** (0.00210)
B+ to B	0.00489 (0.00281)	0.0268*** (0.00383)	0.00375 (0.00495)	-0.0153 (0.00821)	-0.00376* (0.00151)	-0.0164*** (0.00494)
B to B-	0.00525 (0.00283)	0.0289*** (0.00431)	0.00642 (0.00690)	-0.00107 (0.0122)	-0.00684* (0.00298)	-0.0326** (0.00996)
C+ to C	0.00585 (0.00304)	0.0272*** (0.00512)	0.0101 (0.00836)	0.0182 (0.0164)	-0.00988* (0.00481)	-0.0515** (0.0159)
C- & below	0.00546 (0.00286)	0.0224*** (0.00660)	0.0162 (0.00931)	0.0391* (0.0187)	-0.0119 (0.00732)	-0.0714** (0.0224)

Education-starting Population

EDUC Grade	STEM	BE	SSH	OP	EDUC	DROP
A+ to A-	0.0170 (0.00930)	0.00979 (0.00595)	0.0311* (0.0139)	-0.00167 (0.0102)	-0.0528** (0.0201)	-0.00347 (0.00428)
B+ to B	0.0180 (0.00979)	0.0145 (0.00829)	0.0467* (0.0196)	-0.00382 (0.0171)	-0.0612* (0.0290)	-0.0142 (0.0155)
B to B-	0.0145 (0.0102)	0.0288 (0.0159)	0.0615* (0.0251)	-0.00678 (0.0213)	-0.0574 (0.0409)	-0.0406 (0.0390)
C+ to C	0.0107 (0.0111)	0.0238 (0.0264)	0.0658* (0.0269)	-0.000808 (0.0181)	-0.0354 (0.0483)	-0.0642 (0.0587)
C- & below	0.00446 (0.00656)	0.0226 (0.0225)	0.0621* (0.0286)	-0.000186 (0.0106)	-0.0150 (0.0357)	-0.0740 (0.0551)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$