DRAFT: DO NOT CITE WITHOUT **AUTHORS' PERMISSION**

Underrepresented Students Entering STEM Fields

Casey E. George-Jackson[1]
Gregory S. Kienzl
William T. Trent

Educational Policy Studies
University of Illinois at Urbana-Champaign

[1] Corresponding author: cgeo@uiuc.edu

We would like to thank the Sloan Foundation for their generous support of this project and the Mellon Foundation for allowing us access to their data. All errors or omissions are our responsibility.

Abstract

This study will examine the factors that determine the choice of initial college major of first-time women, minorities, and low-income undergraduate students at a large, research-intensive, public university in fall 1999. Of interest is the selection into a STEM (science, technology, engineering, and mathematics) major and, of particular interest, whether these factors are common across all types of STEM and non-STEM majors. The use of an expanded taxonomy of STEM majors demonstrate that underrepresented students participate at higher rates in science-oriented fields rather than more prestigious STEM fields as compared to their participation using a more narrow definition of STEM. Nested logistic regression is used to estimate key factors considered in this study. The results of this study will contribute to institutional effectiveness by identifying possible admittance patterns and/or impediments to educational and diversity goals. Policy improvements are offered based on the results to satisfy both these goals.

Introduction

The STEM (science, technology, engineering and mathematics) education pipeline—from primary education through secondary and postsecondary levels—has been widely studied in an concerted effort to pinpoint periods or circumstances that both preclude and permit students to becoming involved in STEM fields (and eventually STEM occupations). Despite the increased attention, the number of women, non-Asian minorities, and low-income (hereafter, "underrepresented") students in the STEM pipeline has only modestly improved over the past three decades. A considerable disparity still exists between well-represented (i.e., men, whites, Asians, and non-Pell eligible students) and underrepresented populations in these fields (Oakes, 1990; Stake & Mares, 2001).

Seymour (2002) notes that the education pipeline to STEM fields becomes narrower for women and minority students, while white males persist at a higher rate in the same fields. Grandy (1998) and Seymour (2002) argue that the need to diversify STEM fields is an issue of access and equity for populations who have historically marginalized. The breadth and depth of research reflects not only the need for qualified individuals to enter the growing scientific workforce, but also increases in the number of programs designed to expand access to the STEM fields. As shown by Camp (1997), Sax (2000), and Seymour (2002), women's underrepresentation in these fields at the undergraduate level also has long-term, negative reverberations in terms of graduate program enrollment and professional and academic careers in STEM fields, which further erodes and threatens state and national economic competitiveness. The same holds true for minority and low-income students, particularly as low-income students may be deterred from pursuing degrees in STEM fields due to higher costs (i.e., higher tuition and opportunity costs as compared to other majors) and longer time-to-degree, which may delay entry in the labor market.

In short, the diversification of higher education that has occurred due to major societal, programmatic, and policy changes and efforts in the last 40 years has not been fully realized in the nation's top colleges and universities, or in specific fields such as the STEM fields. Minorities and women remain underrepresented in the sciences, even though these populations have increased their representation and participation in other areas of higher education and in the workforce (Rosenbloom, et al., in press). The factors that contribute to the challenges of access for underrepresented students, however, continue to confound educators, researchers, college administrators, and policymakers. While some have argued for interventions in the early and middle grades (Hewson, Kahle, Scantlebury & Davies, 2001), large, public research universities represent a significant conduit in the pipeline and deserve further inspection. Thus, the research question addressed in this study is: What factors influence underrepresented students' entrance into the STEM fields at a large, public, research university?

Literature Review

A number of important background factors affect student's entry into STEM fields. Factors such as high school characteristics, the number and level of science and math courses offered, the number of math and science Advanced Placement (AP) courses available, access to laboratories, types of college preparatory and career course offerings, and resources available to high schools and the communities they serve have all been examined and shown to play an important role. Although mentioned here, we cannot test these factors with the data available for our study. On the other hand, we have information on student's socioeconomic status, parents' level of education, student's academic preparedness for college, and the cost of attendance for pursuing a given college major.[1]

[1] A review of literature on women's departure from the STEM fields highlights a number of plausible explanations which recognize individual, institutional, and societal factors, including differential academic preparation as compared to men, girls' negative experiences with science, the lack of female role models in scientific fields, pedagogy which

These factors, which have also been shown to influence students' choice of major, form the backbone of the current study and are worth further discussion.

Researchers have offered various explanations for the continued underrepresentation—and by extension educational and social inequities—in the STEM fields. Entry into the STEM pipeline at the beginning is of vital importance, as the process of entering a STEM-related occupation starts as early as elementary and secondary school, continues through high school, and into college. Individuals who deviate from the STEM pipeline—by individual choice or other factors—will encounter great difficulty in re-entering the pipeline at a later point in time, or may never return (Elliott, Strenta, Adair, Matier & Scott, 1996). This is due to the fact that preparing to enter into STEM majors and occupations is a sequential and largely linear process given the prerequisites that are required to advance in the sciences and to continue in the STEM pipeline from year-to-year. Declaring a STEM major early in college is critical to minimizing additional time, funds, and opportunity costs spent in pursuing the degree "because the prerequisites for engineering coursework increase the 'cost' to switch into engineering for students who lack these courses" (Frehill, 1997, p. 228).

Although white and Asian males are well-represented in the STEM fields, African Americans are most likely than any other race/ethnicity group to major in science- and math-related fields (Staniec, 2004; Trusty, 2002). In fact, "after controlling for other factors that affect major choice, neither all females nor all minorities are, in fact, underrepresented" (Staniec, 2004, p. 556). In a study of three cohorts of college students at a small liberal arts college, "controlling for these differences [demographics, family influences, academic preparation, academic self-efficacy, and

---

benefits males, and differences in how boys and girls are socialized (Blickenstaff, 2005). Many of these are also applicable to explaining the underrepresentation of racial and ethnic minorities. Socialization of individual's roles in society, role models, the climate of university departments and workplaces, culture of STEM, and even time-context considerations are now accepted as areas worthy of investigation into explaining women and minorities' departure from STEM at various points in the pipeline.

political views], we can see that males and females are just as likely to choose a non-science major over a science major" (Porter & Umbach, 2006, p. 441). This finding offers a reminder that factors affecting the choice of non-STEM majors are just as important to investigate as factors that affect the choice of STEM majors. Important differences exist within the gender categories of who pursues a STEM major. Asian women are more likely than other women to major in STEM fields, as "once Asian women make it to college, they appear to be less affected by the gender-based segregation of college majors" (Song & Glick, 2004, p. 1413). Within men, Asians are most likely to pursue math and science majors compared to other racial and ethnic groups (Trusty, 2002).

Trusty (2002) notes that African American and Hispanic male students from higher socioeconomic backgrounds pursue science and math majors in college at a higher rate than those from lower socioeconomic backgrounds, but females did not display the same differential rate of participation between race and socioeconomic status. Parents' level of education has had inconsistent affect on a student's choice of major in college, with one researcher finding that only African American male students were impacted positively according to their mother's level of education (Maple & Stage, 1991). These are just a few of the many factors investigated in research on differential participation in the STEM fields. What continues to be evident in the dearth of research on reasons for women and minority's underrepresentation in the STEM fields is that the factors and explanations identified thus far "remain unclear and in some cases quite controversial" (Rosenbloom et al., in press, p. 2).

The sociology of education emphasizes the importance of parents' level of education on the educational and occupational attainment of their children, beginning with the status attainment model presented by Blau & Duncan (1967), which investigated the impact of fathers' educational and occupational attainment on sons' educational and occupational attainment. Today, parental factors that impact an individual's choice of major extends beyond parents' level of education.

Parents who are employed in STEM-related fields have a positive impact on their children (male and female) choosing to major in STEM (Leslie, McClure, Oaxaca, 1998), with the greatest impact being on Hispanic males and Black males (Leslie & Oaxaca, 1998). A study of students at historically Black colleges and universities (HBCUs) who persisted in STEM indicated that they were positively impacted in their choice to pursue a STEM major due to their parent's education and careers in scientific fields (Fleming, Engerman & Griffin, 2005).

Staniec (2004) found that females who were raised in single-parent homes were less likely to choose a STEM major than those raised in two-parent households. Students who originated from higher socioeconomic backgrounds were more likely to major in STEM fields, with the largest effects found for Hispanic men. Trusty (2002) notes that African American and Hispanic male students from higher socioeconomic backgrounds pursue science and math majors in college at a higher rate than those from lower socioeconomic backgrounds, but females did not display the same differential rate of participation between race and socioeconomic status. Overall, African American women are about twice as likely as white women to declare a STEM major (Hanson, 2004).

Academic preparation for entrance to a STEM major can be measured by courses taken in high school, performance in those courses (as measured by grades), and performance on standardized college entrance exams (as measured by test scores). A high level of academic preparation on each of these measures is often typical of students in STEM majors as compared to other majors (Levine & Wycokoff, 1991). This can be a function of both admissions requirements to STEM majors, and a student's level of academic preparation may be correlated with choosing a major based on related future career earnings (Song & Glick, 2004). Overall, "taking more science courses in high school, having higher self-ratings relative to mathematics and scientific abilities, and having higher indices of family background not only enhances the likelihood of the selection of a quantitative field of study, but also enhances the likelihood of selecting engineering or the

physical sciences as a major" (Ethington, 2001, p. 359), suggesting that different measures of academic preparation work collectively to increase the probability of declaring a STEM major. A students' choice of STEM major has been found to be positively impacted by the number of courses the student completed in high school in similar disciplines, with the number of math and science courses taken prior to college having a positive impact on whites and African Americans (Maple & Stage, 1991).

Several studies have found that higher preparation for women in math and science results in a greater probability of declaring a STEM major, specifically "taking the most academically intensive math courses—trigonometry, pre-calculus, calculus" (Trusty, 2002, p. 471), and that scoring high on the SAT (Turner & Bowen, 1999). Increasing the level of academic preparation of women has been suggested as a way to increase their participation in these fields in college, however, results from two studies suggest that other interventions will be necessary as well. Frehill (1997) cautions that "increasing the number of math and science courses that girls take in high school would increase the proportion of women who choose engineering by a modest amount" (p. 242), while Turner and Bowen (1999) found that gender "differences in SAT scores account for less than half of the total gender gap" in STEM fields (p. 305). These findings suggest that women who choose to declare non-STEM majors are making choices based on a combination of factors, with academic preparation being only one of many. Men may choose STEM fields at a greater rate than women as a result of "the joint product of variations in SAT scores (higher math scores for the men) and the relatively stronger preferences among men for these fields, though the relative magnitudes of these effects are not at all equal" (Turner & Bowen, 1999, p. 304).

Research Objectives

This project will estimate the effect of individual and institutional factors on underrepresented students' enrollment in STEM fields in comparison to that of their well-represented counterparts. By specifically modeling initial major choice, this study will provide a more accurate picture of the matriculation decisions of underrepresented students in STEM fields during a critical period in postsecondary education. The results of this research project will also contribute to institutional effectiveness by identifying possible interventions that may satisfy both educational and diversity goals. The following research questions will be examined:

1. Based on their pre-college academic qualifications and net price of attendance, do more qualified undergraduate women, minorities, or low-income students opt out of STEM programs in favor of non-STEM programs, as compared to students from traditionally well-represented groups?
2. Do the factors that influence opting in differ within STEM fields, such as between Hard STEM (e.g., engineering, mathematics), Soft STEM (e.g., agricultural sciences), and STEM-Related fields (e.g., biological sciences, psychology)?

Data

The proposed study uses information on a single cohort of first-time freshmen who matriculated to a large, research-intensive university in fall 1999. The data were originally compiled as part of a larger collection of several institutions' data coordinated by the Andrew W. Mellon Foundation, collectively referred to as the *Public University Database*. The cohort used in the analysis consists of over 6,200 first-time, full-time, domestic freshman in fall 1999[2,3,4] Information

[2] Students who began their college studies in summer 1999 were excluded. Of the 77 students who began their studies in summer 1999, only 7 switched programs between summer 1999 and fall 1999.

on student's social backgrounds, pre-college academic qualifications, academic performance during college, and type and amount of financial aid received are also provided in the dataset. Full information was available to 4,880 students, which is the sample size used in the nested logistic regression analysis.

Methodology

A combination of quantitative methods and statistical models will be used in an attempt to answer the research questions posed in this study. The first research question—who opts into STEM fields—will be addressed using both descriptive statistics and nested logistic regression (i.e., conditional decision-tree) models with socio-demographic characteristics, pre-college academic qualifications as measured by SAT scores,[5] and financial attributes (e.g., amount of financial aid received) included as explanatory variables. Majors were further categorized as Hard STEM (e.g., engineering and computer science), Soft STEM (e.g., agricultural sciences), STEM-Related (e.g., health-related fields), and Non-STEM (e.g., humanities and fine arts) fields and the models rerun to determine if who opts into each fields differs between these categories. Details of the estimation strategy are provided below.

In order to compare various science-oriented fields, rather than simply focus on high-status STEM majors (e.g., engineering, computer science), a four-category taxonomy was used to differentiate between different categories of majors (see Chart 1). The first two-digits of CIP codes for students' initial major were used to create the system.

---

[3] Twenty one observations were lost by excluding part-time students. Federal and institutional measures of full-time and part-time status agreed with each other in the dataset.

[4] Although the dataset only includes students who enrolled in fall 1999, there were nearly 18,000 applicants and over 12,500 students were accepted to the university in fall 1999. Including transfer students, nearly 6,500 new students enrolled in fall 1999.

[5] ACT scores were converted to SAT scores to allow for ease of comparison of students' standardized test scores. Other measures of pre-college academic qualifications such as high school GPA or specific courses taken in high school are not included in the dataset.

Chart 1. Classification of college major by CIP codes

Category	2-digit CIP Code	CIP Family
Hard STEM	11	Computer and Information Sciences and Support Services
	14	Engineering
	27	Mathematics and Statistics
	40	Physical Sciences
Soft STEM	01	Agriculture, Agriculture Operations, and Related Services
	03	Natural Resources and Conservation
	04	Architecture and Related Services
STEM-Related	26	Biological and Biomedical Sciences
	42	Psychology
	51	Health Professions and Related Clinical Sciences
Non-STEM	09	Communication, Journalism, and Related Programs
	13	Education
	16	Foreign Languages, Literatures, and Linguistics
	19	Family and Consumer Sciences/Human Sciences
	23	English Language and Literature/Letters
	24	Liberal Arts and Sciences, General Studies and Humanities
	30	Multi/Interdisciplinary Studies
	31	Parks, Recreation, Leisure, and Fitness Studies
	38	Philosophy and Religious Studies
	45	Social Sciences
	49	Transportation and Materials Moving
	50	Visual and Performing Arts
	52	Business, Management, Marketing, and Related Support Services
	54	History (New): Study and Interpretation of Past Events, Institutions, Issues, and Cultures
	99[6]	General Curriculum

[6] CIP code "99" was created by the authors for students enrolled in "General Curriculum" as they do not belong to a specific college or school. These students are distinctive than those enrolled in "General Studies" listed in CIP code "24".

Estimation Strategy

To estimate students' likelihood of entering a STEM major from a set of alternative majors, a nested logistic regression model will be used. The nested logistic regression requires the data to be organized as pairwise combinations of students, i, and unique combinations of majors, J. For the sake of simplicity, the initial major choice set is restricted to four general programs of study: Hard STEM, Soft STEM, STEM-Related and Non-STEM. Nested logistic regression models relax the assumption of independently distributed errors and the independence of irrelevant alternatives inherent in conditional logistic regression models by clustering selected alternatives into hierarchical "nests" (Greene, 2008). For this study, the four general programs of study comprise the first nest at the top of hierarchy, and specific majors that share common attributes (tuition and fee charge by resident status) comprise the second (or bottom) nest. In all, there are 14 unique combinations of majors.[7]

The estimation is based on students choosing major j at institution Q based on their individual characteristics, which are contained in vector X and attributes of the institution. These factors affect students' demand for one of the four majors listed above. Students select a particular major in their first year that maximizes their utility, $U(Y_j, X_i)$. Estimating this utility is done via a linear function of student characteristics

$$U(Q_j, X_i) = \beta_1 Z_{ij1} + \beta_2 Z_{ij2} + \ldots + \beta_k Z_{ijk} + \varepsilon_{ij}$$

where β is a vector of parameters, Z_{ij} are the variables that affect utility, k is the total number of variables in the model, and ε_{ij} is idiosyncratic error. Z may include variables that describe the elements of Y or interact with Y and X to form match-specific measures.

[7] The distribution of the 14 combinations are shown in Table 10.

Discussion of Results

The results from simple descriptive analysis are shown in Tables 1 to 9, provided in Appendix A. The results are discussed below.

Social background

Table 1 offers an overview of the demographic and social background characteristics of the students in the dataset. Of the 6,212 students who entered in the university in fall 1999, 51 percent were male and 49 percent were female. In terms of the racial and ethnic composition of the students, 72 percent were White, 13 percent were Asian, 7 percent were Black, 6 percent were Hispanic, less than one percent were Native American, and one percent of the students' race or ethnicity was unknown. Male and female representation within racial and ethnic categories is approximately equal. The vast majority (93 percent) of students are in-state residents. Of the students who reported their parent's highest level of education (n=4,480), 50 percent of fathers had a college degree or higher and 46 percent of mothers had a college degree or higher.[8]

Academic preparation as measured by SAT scores

Information available on students' academic preparation prior to entering college is limited to their standardized test scores (see Table 2). The mean SAT Math score of the students was 629, the mean SAT Verbal score was 598, and the mean total SAT score was 1227. Comparing the scores of students enrolled in each of the four categories of majors, students enrolled in the Hard STEM fields had the highest mean SAT score (1329), followed by those majoring in STEM-Related fields (1235), Non-STEM fields (1189) and Soft STEM fields (1179). Within each of the four categories of majors, the mean of SAT total scores by males was higher than the mean of SAT scores by females, varying between a 15 point difference in the Soft STEM fields to a 51 point difference in the STEM-Related fields. Across all four major categories, males had higher mean

[8] Nearly 30 percent of father's and mother's highest level of education each is missing from the dataset.

SAT Math scores, while females had higher mean SAT Verbal scores in the Soft STEM and Non-STEM majors. Low-income students, as indicated by those who received a Pell grant, had lower Math, Verbal, and Total SAT scores across all four categories of majors, compared to students who filed FAFSA but did not qualify for a Pell grant (see Table 3).

Examining differences in SAT scores by race and ethnicity and according to student's initial major provides additional insights into this cohort of undergraduate students (see Table 4). Asian males who entered the Hard STEM fields had the highest mean SAT Math score (718), while those with an unknown race/ethnicity entering the Hard STEM fields had the highest mean SAT Verbal score (664) and the highest mean Total SAT score (1372). Within each race/ethnic category, students entering the Hard STEM fields had the highest mean Math, Verbal, and Total SAT scores as compared to other majors. Students entering the Non-STEM fields had the lowest mean Total SAT scores within each race/ethnic category except for Asians and Whites whose lowest mean total SAT scores were of students who entered Soft STEM fields. SAT scores by race/ethnicity and gender were examined next. Within each of the four categories, Blacks had the lowest mean total SAT scores as compared to the other racial/ethnic groups.

Gender differences of standardized test scores within racial and ethnic categories by classification of initial major are shown in Table 5. The mean Total SAT score was highest for each gender within each racial/ethnic category for those who entered the Hard STEM fields, except for Black males (Soft STEM), Hispanic females (Soft STEM), and female students' whose race and ethnicity is unknown (STEM-related). Asian males and females in the Hard STEM fields had the highest mean SAT Math score—725 and 698, respectively. White females and Males of unknown racial/ethnic backgrounds had the highest mean SAT Verbal scores (641 and 669, respectively). Asian males and White males had higher mean SAT Total scores as compared to their female

counterparts, regardless of the major category. However, there was some variation by gender within Blacks, Hispanics, and Native Americans across each of the four major categories.

First major declared

Students' major field of choice the first semester of college (fall 1999) was examined (see Table 6). Overall, 24 percent of students began their undergraduate studies in the Hard STEM fields, 10 percent in the Soft STEM fields, 12 percent in the STEM-Related fields, and 54 percent in Non-STEM fields. While women's underrepresentation in the high-profile, high-status STEM fields such as engineering is well-documented, using this expanded taxonomy of majors reveals a different picture in terms of women's participation in science-oriented majors at the undergraduate level. If only male (37 percent) and female (11 percent) participation in the Hard STEM fields are compared, it appears that very few women initially enroll in the sciences. However, by also investigating enrollment in the Soft STEM and STEM-Related fields, more females are enrolled in each of these categories as compared to males (10 percent versus 9 percent for Soft STEM fields, and 16 percent versus 7 percent for STEM-Related, respectively). Combining these three STEM categories—all of which are science-oriented although to varying degrees—53 percent of men and 38 percent of women are enrolled in STEM majors. Although women are still participating in these fields at lower rates than men, the overall participation of women in science-oriented fields (by including a broader and more nuanced definition of STEM) increases.

The differentiation of STEM fields according to this classification also reveals increase participation for Blacks and Hispanics. If only Hard STEM fields are examined, only 13 percent of Blacks and 17 percent of Hispanics enroll in these majors. Using the expanded taxonomy, these numbers increase to 29 percent and 34 percent, respectively. The increase is also present for Black, Hispanic, and white females. While only 9 percent of Black females, 6 percent of Hispanic females, and 11 percent of white females enrolled in Hard STEM fields, 31 percent of Black females, 29

percent of Hispanic females, and 39 percent of white females are enrolled across the three STEM categories.

Financial aid

Financial aid information is available for students who filed a FAFSA form, which is 70 percent of the students in the study (see Table 7). Of those who completed a FAFSA form, 15 percent received a Pell grant, signifying low-income status. The net price, which is calculated from the tuition and fee charges for the first year of attendance minus any aid, averaged approximately $10,500 for Hard STEM majors, $9,500 for Soft STEM majors, $9,900 for STEM-Related majors, and $9,700 for STEM-Related majors (see Table 8). Women in the Hard and Soft STEM fields have slightly lower net prices as compared to their male counterparts. Asians in the Hard STEM fields have the highest average net prices of all racial and ethnic groups ($11,300), nearly double the net price of Blacks and Hispanics ($6,600 each) enrolled in the same majors. Black females enrolled in the Soft STEM fields have the lowest average net price of any female across each of the four categories ($2,000) followed by Hispanic females enrolled in the Hard STEM fields ($4,800). As expected, the net price of students who received Pell grants is considerably lower than those who did not receive Pell grants, and the net price of out-of-state students is higher than in-state students, across all four categories. Information on the expected family contribution (EFC) is also available for students who filed for FAFSA (see Table 9). Overall, students in the Hard STEM fields have the highest EFC ($16,900). Excluding students of unknown race and ethnicity, white students across all four categories of majors have the highest EFC. Again, as expected, the EFC of students who received Pell grants was considerably lower than those who did not receive Pell grants, with the average EFC of Pell recipients less than $1,200 across each of the field categories.

Nested logistic regression

Drawing from the descriptive analysis, the nested logistic regression allows for the key factors mentioned above to be estimated. This analysis addresses the second research question; namely, are the factors that affect opting into STEM consistent across other majors? The nested logistic regression results generally confirm the earlier findings, but also highlight some important differences that the descriptive analysis was unable to detect.

As noted earlier, females and non-Asian minorities are generally less likely to enroll in STEM than their white and Asian male counterparts.[9] The results from the nested logistic regression, however, reveal a more complex picture. At this large, research university, for example, females are much less likely to enroll in Hard STEM majors than their male peers (see Table 11). This effect—while still negative and significant—is much smaller for Soft STEM majors, indicating that the barriers to entry might be less for these types of majors. With regard to STEM-Related majors, females are more likely than males to enroll, *ceteris paribus*. These findings suggest that the STEM pipeline is not entirely closed to females, but instead may be directed toward less prestigious science majors.

The predicted probabilities of selecting one of the 14 alternatives from the choice set are shown in Table 10. Shaded rows are for out-of-state students and those indexed with a B cost slightly more to attend than those indexed with an A. Cost is measured by subtracting the average aid awarded to those in the major category from the tuition/fees and other related charges. At the first (or upper) level, the model fits the data quite well. For instance, 25.9 percent of first-time, full-time undergraduates enrolled in a Hard STEM major. The nested logistic regression predicted 25.7 percent of these students to do so. The predicted probabilities from the selected socio-demographic characteristics also closely mirror their actual distributions at the first level. The predicted

[9] Note that due to the small sample size, Native Americans (n=12) were merged into the "Other Race" category for the purpose of running the model, along with students with unknown race and ethnicity..

probabilities align less well with the actual distribution at the second level. Although not the primary focus of this study, preliminary examinations of this finding suggest that the model overestimates the "costlier" Hard STEM alternatives but underestimates these alternatives for STEM-Related and Non-STEM majors. In other words, students at this institution are more willing to enroll in and pay for the higher price Hard STEM majors due to the expected economic return on their investment.

Compared to first-time, full-time white students, similar Black and Hispanic students are much less likely to enroll in Soft STEM majors than Hard STEM, STEM-Related or Non-STEM majors. Despite the relative advantage to Soft STEM, the barrier to Hard STEM majors still exists for Black and Hispanic students. Future analysis will explore the decision-making of female Black and Hispanic students. There were no statistical differences in the choice of initial major between Pell grant recipients and non-recipients.

Policy and Program Implications

This study makes several contributions. First, it contributes to the collective understanding of representation of women and minorities in STEM fields at the undergraduate level, as well as factors that affect major field choice. A better understanding of the key factors that affect underrepresented students' matriculation in the STEM fields, such as SAT score of the average freshmen and the tuition and fees charged as demonstrated by this study, will help policymakers and postsecondary administrators design programs to promote enrollment so that students can benefit more fully from entering undergraduate programs that are associated with earning higher incomes, experiencing greater social mobility, and benefiting from enhanced professional opportunities. As May and Chubin (2003) contend that the previous underrepresentation by minorities in the STEM

fields is a repairable problem through program and policy improvements, the research findings presented here can be used to inform institutional-based recruitment interventions.

In addition, the taxonomy used in this research that includes additional science-oriented majors that are traditionally excluded from the literature on STEM fields demonstrates the importance of using a more inclusive definition of what compromises "STEM." Although women, non-Asian minorities, and low-income students remain underrepresented as compared to their white and Asian male counterparts, their participation rates in the sciences increases when Soft STEM and STEM-Related majors are included in the measure of STEM participation. This finding challenges notions of disinterest or disengagement by underrepresented students in scientific fields at the undergraduate level.

This broad definition of STEM can be used by universities in their recruitment efforts of underrepresented students. By emphasizing participation across a broader swath of scientific programs on a particular campus instead of just focusing on the high-prestige STEM fields, prospective students may be more encouraged to attend the institution. In this sense, universities should examine their scientific community across all science and math-based majors, and highlight the participation rates and achievements of underrepresented students in those fields. Although the finding of increased participation in the sciences is encouraging, research should continue to investigate reasons for underrepresentation in the Hard STEM fields, as well as reasons for participation in the Soft STEM and STEM-Related fields by members of these groups.

The results also inform future financial aid policies as a recruitment device for underrepresented students. Underrepresented students are more likely to enter higher-cost STEM fields if they have financial aid. Universities can target financial aid to these students as a way to reduce potential cost barriers to participating in higher-cost STEM fields.

Lastly, the study also makes a methodological contribution through the use of a nested logistic regression. The nested framework is uniquely suited to handle interrelated choices and allow a relaxation of the IIA (Independence of Irrelevant Alternatives) property.

Limitations and Future Research

A number of limitations exist in this study. First, the data are drawn from a single cohort of students attending a single institution, which limits the generalizability of these findings. Selection bias is present in the dataset due to the data comprising only of students who attended the university and does not include data on students who applied or who were admitted but did not enroll. In addition, students may be self-selecting into specific majors based on interests, motivations, and orientations to future careers; all of which are unobservable in the dataset. The data also lacks variables that could provide more insight into students' pre-college academic qualifications such as their high school grade point average, high school transcript information, or non-cognitive measures. In addition, the use of standardized test scores has somewhat limited utility as an indicator of academic preparation. Furthermore, the university's admission process considers other factors that are not included in the dataset, such as high school coursework requirements, application essays, extracurricular activities, service, and work experience.[10] Finally, important differences within gender and within racial and ethnic groups cannot be examined due to the number of observations of these groups in the dataset.

Some of these limitations will be addressed in future research. The next phase of the research project will expand the model used here to include other large, public, research universities, which will expand the number of institutions included in the study, as well as the total number of observations. The increased number of observations will provide the opportunity to

[10] Applications to specific majors such as those classified as visual and performing arts require portfolios and/or auditions.

disaggregate gender and racial/ethnic groups, allowing for important within groups differences to be examined. Data will also be merged in from the U.S. Department of Education's Common Core Data (CCD) and the Office of Civil Rights (OCR) to provide information on the high schools students' attended, including the number of AP math and science courses offered.[11]

Given that this study only examines the initial major field of choice for students, it is also pertinent to examine persistence and degree attainment in future studies. The dataset follows students for a period of six years, providing semester-by-semester major information, allowing for persistence and completion patterns to be investigated. Of particular interest will also be following students who switch in and out of the various STEM majors and investigating the factors that contribute to these movements.

Lastly, it is important to note that appearing alongside the call for increased educational equity and access to postsecondary education for underrepresented students is the call to strengthen our nation's globally competitive workforce. U.S. Education Secretary Margaret Spellings' Commission on the Future of Higher Education, which released a report based on a year-long review of the status of U.S. higher education in fall 2006, argues for the need for more American-born scientists in order to strengthen the nation's workforce and global competitiveness (U.S. Department of Education, 2006). Thus, in response to this appeal, and to better understand the preparation of a skilled workforce, it is important to also consider the persistence and attainment patterns of those who enter the STEM education pipeline and the impact (and responsibility) of large, research-intensive, public universities in educating undergraduates in these fields. Persistence and degree attainment in the STEM fields will be examined in future studies to determine the educational outcomes of students according to their initial major choice.

[11] Merging these datasets will reduce the number of observations given that the OCR data is based on a stratified random sample of public schools.

References

Blau, P.M., & Duncan, O.D. (1967). *The American occupational structure*. New York: Wiley.

Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? *Gender & Education, 17*(4), 369-386.

Camp, T. (1997). The incredible shrinking pipeline. *Communications of the ACM*, *40*(10), 103-110.

Elliott, R, Strenta, A.C., Adair, R., Matier, M., & Scott, J. (1996). The role of ethnicity in choosing and leaving science in highly selective institutions. *Research in Higher Education, 37*(6), 681-709.

Fleming, L., Engerman, K. & Griffin, A. (2005). *Persistence in engineering education: Experiences of first year students at a historically black university*. Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition. June 12-15, 2005, Portland, OR. Retrieved July 2, 2008, from http://www.asee.org/acPapers/2005-1786_Final.pdf

Frehill, L. M. (1997). Education and occupational sex segregation: The Decision to Major in Engineering. *The Sociological Quarterly, 38*(2), 225–249.

Grandy, J. (1998). Persistence in science of high ability minority students. *The Journal of Higher Education, 69*(6), 589-620.

Greene, W.H. (2008). *Econometric analysis* (6th ed.). MacMillan: New York.

Hanson, S.L. (2004). African American women in science: Experiences from high school through the post-secondary years and beyond. *National Women's Study Association Journal, 16*(1), 96-115.

Hewson, P.W., Kahle, J.B., Scantlebury, K. & Davies, D. (2001). Equitable science education in urban middle schools: Do reform efforts make a difference? *Journal of Research in Science Teaching*, *38*(10), 1130–1144.

Leslie, L. L., McClure, G. T., & Oaxaca, R. L. (1998). Women and Minorities in Science and Engineering: A Life Sequence Analysis. *The Journal of Higher Education, 69*(3), 239-276.

Leslie, L.L., & Oaxaca, R.L. (1998). Women and minorities in higher education. In J.C. Smart (Ed.). , *Higher education: Handbook of theory and research*, Vol. 13, New York: Agathon Press, 304-352.

Levine, J., & Wycokoff, J. (1991). Predicting Persistence and Success in Baccalaureate Engineering. *Education, 111*(4), 461-468.

Maple, S.A., & Stage, F.K. (1991). Influences on the choice of math/science major by gender and ethnicity. *American Educational Research Journal, 28*(1), 37-60.

May, G.S & Chubin, D.E. (2003) A retrospective on undergraduate engineering success for underrepresented minority students. *Journal of Engineering Education, 92*(1), 27-40.

Oakes, J. (1990). Opportunities, achievement, and choice: Women and minority students in science and mathematics. *Review of Research in Education*, *16*, 153–222.

Porter, S. R., & Umbach, P. D. (2006). College major choice: An analysis of person–environment fit. *Research in Higher Education, 47*(4), 429-449.

Rosenbloom, J.L., Ash, R.A., Dupont, B., & Coder, L. (in press). *Why are there so few women in information technology? Assessing the role of personality in career choices*. Journal of Economic Psychology.

Sax, L.J. (2000). Undergraduate science majors: Gender differences in who goes to graduate school. *The Review of Higher Education*, 24(2), 153–172.

Seymour, E. (2002). Tracking the processes of change in U.S. undergraduate education in science, mathematics, engineering, and technology. *Science Education*, *86*, 79–105.

Song, C., & Glick, J. E. (2004). College Attendance and Choice of College Majors Among Asian-American Students. *Social Science Quarterly (Blackwell Publishing Limited), 85*(5), 1401-1421.

Stake, J.E. & Mares, K.R. (2001). Science enrichment programs for gifted high school girls and boys: Predictors of program impact on science confidence and motivation. *Journal of Research in Science Teaching*, *38*(10), 1065–1088.

Staniec, J. F. O. (2004). The effects of race, sex, and the expected returns on the choice of college major. *Eastern Economic Journal, 30*(4), 549-562.

Trusty, J. (2002). Effects of High School Course-Taking and Other Variables on Choice of Science and Mathematics College Majors. *Journal of Counseling & Development, 80*(4), 464.

Turner, S. E., & Bowen, W. G. (1999). Choice of Major: The Changing (Unchanging) Gender Gap. *Industrial and Labor Relations Review, 52*(2), 289-313.

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: Author. Retrieved September 28, 2006, from http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf.

Appendix A

Table 1. Demographic and social background information, 1999

Gender	Male	3193	51.4%
	Female	3019	48.6%
Race and Ethnicity	Total	6212	100.0%
	Asian	812	13.1%
	Black	459	7.4%
	Hispanic	378	6.1%
	Native American	12	0.2%
	White	4461	71.8%
	Unknown	90	1.4%
Residency Status	Out-of-State	442	7.1%
	In-State	5770	92.9%
Father's Highest Level of Education	At least high school	1411	22.7%
	College and higher	3086	49.7%
	Missing	1715	27.6%
Mother's Highest Level of Education	At least high school	1653	26.6%
	College and higher	2877	46.3%
	Missing	1682	27.1%

Source: Public University Database (2008). Author's calculations.

Table 2. Mean SAT scores of initial major by gender, 1999

		Hard STEM	Soft STEM	STEM-Related	Non-STEM
Male	Math	705	621	657	620
	Verbal	633	566	613	581
	Total	1338	1187	1270	1201
Female	Math	673	590	613	589
	Verbal	625	582	606	590
	Total	1298	1172	1219	1179
Total	Math	698	605	626	603
	Verbal	631	574	609	586
	Total	1329	1179	1235	1189

Source: Public University Database (2008). Author's calculations.

Table 3. Mean SAT score of initial major by Pell status, 1999

		Hard STEM	Soft STEM	STEM-Related	Non-STEM
Did not Receive Pell	Math	703	609	634	612
	Verbal	637	578	616	595
	Total	1340	1187	1250	1207
Received Pell	Math	662	577	582	556
	Verbal	592	552	567	539
	Total	1255	1129	1150	1095

Source: Public University Database (2008). Author's calculations.

Table 4. Mean SAT scores of initial major by race and ethnicity, 1999

		Hard STEM	Soft STEM	STEM-Related	Non-STEM
Asian	Math	718	634	662	649
	Verbal	623	561	609	580
	Total	1342	1196	1271	1229
Black	Math	593	528	526	484
	Verbal	548	539	539	492
	Total	1141	1067	1064	976
Hispanic	Math	630	606	571	547
	Verbal	567	561	561	543
	Total	1197	1167	1132	1090
Native American	Math	650	600	N/A	586
	Verbal	550	550	N/A	539
	Total	1200	1150	N/A	1124
White	Math	702	604	636	618
	Verbal	641	577	622	603
	Total	1343	1181	1258	1221
Unknown	Math	708	652	648	622
	Verbal	664	606	627	609
	Total	1372	1258	1274	1230
Total	Math	698	605	626	603
	Verbal	631	574	609	586
	Total	1329	1179	1235	1189

Source: Public University Database (2008). Author's calculations.

Table 5. Mean SAT scores of initial major by selected socio-demographic characteristics, 1999

			Hard STEM	Soft STEM	STEM-Related	Non-STEM
Asian	Male	Math	725	642	677	666
		Verbal	628	556	616	584
		Total	1353	1198	1292	1250
	Female	Math	698	627	650	635
		Verbal	609	566	604	577
		Total	1307	1193	1254	1213
Black	Male	Math	607	606	544	503
		Verbal	545	566	514	490
		Total	1152	1172	1058	992
	Female	Math	575	492	520	469
		Verbal	552	527	546	494
		Total	1126	1019	1066	963
Hispanic	Male	Math	633	596	594	560
		Verbal	567	553	542	540
		Total	1200	1149	1135	1099
	Female	Math	618	640	565	536
		Verbal	567	593	566	545
		Total	1184	1233	1131	1082
Native American	Male	Math	650	600	N/A	583
		Verbal	550	550	N/A	523
		Total	1200	1150	N/A	1107
	Female	Math	N/A	N/A	N/A	590
		Verbal	N/A	N/A	N/A	570
		Total	N/A	N/A	N/A	1160
White	Male	Math	708	619	665	636
		Verbal	641	567	627	597
		Total	1349	1186	1292	1233
	Female	Math	678	590	623	603
		Verbal	641	585	619	608
		Total	1319	1175	1243	1211
Unknown	Male	Math	710	713	670	630
		Verbal	669	627	630	580
		Total	1379	1340	1300	1210
	Female	Math	680	560	640	612
		Verbal	550	575	626	641
		Total	1230	1135	1266	1253
Total	Male	Math	705	621	657	620
		Verbal	633	566	613	581
		Total	1338	1187	1270	1201
	Female	Math	673	590	613	589
		Verbal	625	582	606	590
		Total	1298	1172	1219	1179

Source: Public University Database (2008). Author's calculations.

Table 6. Initial major by selected socio-demographic characteristics, 1999

		Overall					Women Only				
		Hard STEM	Soft STEM	STEM-Related	Non-STEM	Total	Hard STEM	Soft STEM	STEM-Related	Non-STEM	Total
Overall		1521	607	717	3367	6212					
		24%	10%	12%	54%	100%					
Male		1177	296	224	1496	3193					
	Row %	37%	9%	7%	47%	100%					
	Column %	77%	49%	7%	47%	51%					
Female		344	311	493	1871	3019					
	Row %	11%	10%	16%	62%	100%					
	Column %	23%	51%	69%	56%	49%					
Asian		295	49	112	356	812	75	25	63	202	365
	Row %	36%	6%	14%	44%	100%	21%	7%	17%	55%	100%
	Column %	19%	8%	16%	11%	13%	22%	8%	13%	11%	12%
Black		59	16	61	323	459	25	11	47	184	267
	Row %	13%	3%	13%	70%	100%	9%	4%	18%	69%	100%
	Column %	4%	3%	9%	10%	7%	7%	4%	10%	10%	9%
Hispanic		65	14	49	250	378	12	3	38	132	185
	Row %	17%	4%	13%	66%	100%	6%	2%	21%	71%	100%
	Column %	4%	2%	7%	7%	6%	3%	1%	8%	7%	6%
Native American		2	1	0	9	12	2	1	0	6	9
	Row %	17%	8%	0%	75%	100%	0%	0%	0%	100%	100%
	Column %	0.1%	0.2%	0.0%	0.3%	0.2%	0.0%	0.0%	0.0%	0.2%	0.1%
White		1077	522	483	2379	4461	231	270	336	1,326	2,163
	Row %	24%	12%	11%	53%	100%	11%	12%	16%	61%	100%
	Column %	71%	86%	67%	71%	72%	67%	87%	68%	71%	72%
Unknown		23	5	12	50	90	1	2	9	24	36
	Row %	26%	6%	13%	56%	100%	3%	6%	25%	67%	100%
	Column %	2%	1%	2%	1%	1%	0%	1%	2%	1%	1%
Non-Resident		194	19	38	191	442					
	Row %	44%	4%	9%	43%	100%					
	Column %	13%	3%	5%	6%	7%					
Resident		1327	588	679	3176	5770					
	Row %	23%	10%	12%	55%	100%					
	Column %	87%	97%	95%	94%	93%					

Source: Public University Database (2008).
Author's calculations.

Table 7. Financial aid information of initial major by selected socio-demographic characteristics, 1999

	Overall				
	Hard STEM	Soft STEM	STEM-Related	Non-STEM	Total
Did not Complete FAFSA	399	188	205	1085	1877
	21%	10%	11%	58%	100%
	26%	31%	29%	32%	30%
Completed FAFSA	1122	419	512	2282	4335
	26%	10%	12%	53%	100%
	74%	69%	71%	68%	70%
Did not Receive Pell	1324	527	608	2829	5288
	25%	10%	12%	54%	100%
	87%	87%	85%	84%	85%
Received Pell	197	80	109	538	924
	21%	9%	12%	58%	100%
	13%	13%	15%	16%	15%

Source: Public University Database (2008). Author's calculations.

Table 8. Mean net price of initial major by selected socio-demographic characteristics, 1999

	Overall				Women Only			
	Hard STEM	Soft STEM	STEM-Related	Non-STEM	Hard STEM	Soft STEM	STEM-Related	Non-STEM
Overall	$10,542	$9,573	$9,911	$9,790				
Male	$10,738	$9,843	$9,788	$9,881				
Female	$9,874	$9,316	$9,967	$9,718				
Asian	$11,304	$9,851	$10,182	$9,585	$9,740	$9,609	$10,081	$9,340
Black	$6,663	$2,706	$6,633	$5,233	$6,924	$2,017	$7,006	$5,061
Hispanic	$6,667	$7,034	$7,199	$7,136	$4,841	$5,042	$6,765	$7,099
Native American	$10,690	$15,011	N/A	$9,197	N/A	N/A	N/A	$8,874
White	$10,764	$9,805	$10,478	$10,697	$10,481	$9,605	$10,633	$10,675
Unknown	$11,291	$10,648	$12,317	$10,938	$13,846	$13,270	$13,265	$10,265
Did not Receive Pell	$11,522	$10,559	$11,065	$11,013				
Received Pell	$3,961	$3,074	$3,472	$3,362				
Out-of-State	$16,767	$15,499	$17,318	$16,121				
In-State	$9,632	$9,381	$9,497	$9,410				

Source: Public University Database (2008). Author's calculations.

Table 9. Mean expected family contribution of initial major by selected socio-demographic characteristics, 1999

	Overall				Women Only			
	Hard STEM	Soft STEM	STEM-Related	Non-STEM	Hard STEM	Soft STEM	STEM-Related	Non-STEM
Overall	$16,960	$14,822	$13,952	$14,041				
Male	$17,030	$14,888	$15,318	$14,153				
Female	$16,745	$14,763	$13,309	$13,955				
Asian	$14,097	$12,175	$14,044	$9,899	$13,034	$9,574	$14,484	$8,936
Black	$8,078	$4,511	$7,919	$6,145	$10,673	$3,032	$7,303	$5,528
Hispanic	$9,798	$7,719	$10,208	$8,917	$6,906	$10,416	$9,425	$9,145
Native American	$0	$21,205	N/A	$13,681	N/A	N/A	N/A	$19,633
White	$18,841	$15,735	$15,360	$16,949	$19,466	$15,911	$14,867	$17,082
Unknown	$20,181	$16,638	$18,162	$19,146	.	$22,902	$11,895	$17,901
Did not Receive Pell	$20,374	$18,065	$17,514	$18,115				
Received Pell	$1,107	$1,121	$1,077	$926				
Out-of-State	$22,476	$14,724	$15,854	$18,088				
In-State	$16,198	$14,825	$13,861	$13,843				

Source: Public University Database (2008). Author's calculations.

Table 10. Comparison between actual and predicted choice of initial major by selected socio-demographic characteristics, 1999

Major	Actual		Predicted		Predicted by selected socio-demographic characteristics						
categories	N	Percent	N	Percent	Male	Female	White	Black	Hispanic	Asian	Pell grant
Hard STEM A	177	4.0%	502	11.2%	16.6%	5.6%	11.3%	5.7%	8.0%	16.1%	9.6%
Hard STEM B	839	18.7%	408	9.1%	13.5%	4.6%	9.2%	4.6%	6.5%	13.1%	7.8%
Hard STEM A	30	0.7%	154	3.4%	5.1%	1.7%	3.5%	1.8%	2.5%	5.0%	2.9%
Hard STEM B	116	2.6%	86	1.9%	2.8%	1.0%	1.9%	1.0%	1.4%	2.8%	1.6%
Soft STEM	428	9.6%	442	9.9%	9.6%	10.2%	12.1%	3.7%	4.2%	5.8%	8.5%
Soft STEM	15	0.3%	15	0.3%	0.3%	0.4%	0.4%	0.1%	0.1%	0.2%	0.3%
STEM related A	251	5.6%	211	4.7%	3.4%	6.1%	4.6%	5.0%	4.9%	4.8%	4.4%
STEM related B	250	5.6%	326	7.3%	5.3%	9.4%	7.2%	7.8%	7.6%	7.4%	6.8%
STEM related A	8	0.2%	19	0.4%	0.3%	0.5%	0.4%	0.4%	0.4%	0.4%	0.4%
STEM related B	18	0.4%	15	0.3%	0.2%	0.4%	0.3%	0.3%	0.3%	0.3%	0.3%
Non STEM A	2106	47.0%	1547	34.5%	28.8%	40.4%	32.9%	46.7%	43.1%	29.7%	38.5%
Non STEM B	130	2.9%	733	16.4%	13.7%	19.2%	15.6%	22.1%	20.4%	14.1%	18.3%
Non STEM A	95	2.1%	9	0.2%	0.2%	0.2%	0.2%	0.3%	0.3%	0.2%	0.2%
Non STEM B	17	0.4%	12	0.3%	0.2%	0.3%	0.3%	0.4%	0.3%	0.2%	0.3%

Note: Shaded rows are the results for out-of-state students. Major categories B charge slightly higher tuition than A.

Source: Public University Database (2008). Author's calculations.

Table 11. Nested logit regression results of initial major choice

Second level equation

	Major choice [14 alternatives]
R's SAT score - Avg. SAT score	**-0.373**
	(0.067)
Net price	**-0.142**
	(0.027)

First level equation [Ref: *Non-STEM major*]

	Hard STEM	Soft STEM	STEM Related
Female	**-1.381**	**-0.258**	**0.289**
	(0.082)	(0.102)	(0.095)
Black	**-0.931**	**-1.583**	-0.263
	(0.167)	(0.277)	(0.168)
Hispanic	**-0.540**	**-1.404**	-0.192
	(0.163)	(0.285)	(0.181)
Asian	**0.366**	**-0.670**	0.071
	(0.111)	(0.192)	(0.147)
Other race	-0.245	-0.712	-0.198
	(0.296)	(0.48)	(0.398)
Received Pell grant	-0.106	-0.010	-0.247
	(0.105)	(0.142)	(0.128)
Residency	**-1.217**	**-0.927**	**-1.629**
	(0.147)	(0.108)	(0.11)
Father's education level:	-0.074	**-0.294**	**-0.388**
At least some college	(0.096)	(0.119)	(0.108)
Mother's education level:	**0.265**	-0.052	0.002
At least some college	(0.092)	(0.116)	(0.106)
Number of observations		62720	
Number of cases		4480	
Log likelihood		-8490.96	

Note: Statistically significant (p<0.05) are shown in bold.
Source: Public University Database (2008). Author's calculations.