2017

Analysis of the Effects of First Year Advisors and First Year Mentors on a Wellesley Student's Choice of STEM vs. non STEM Major

Kelly Kung
kkung@wellesley.edu

Analysis of the Effects of First Year Advisors and First Year Mentors on a Wellesley Student's Choice of STEM vs. non STEM Major

Kelly Kung

Submitted in Partial Fulfillment of the Prerequisite for Honors in the Mathematics Department

April 2017

Acknowledgements

I would like to take this time to thank everyone who made this thesis possible. I would like to thank Pamela Taylor and the Office of Institutional Research for providing me with the data used for this thesis. I would also like to thank Dean John O'Keefe and Dean Lori Tenser for providing additional data on mentor groups and advising pairs and meeting with me to explain the process of assigning mentors and advisors. Additionally, Ravi Ravishanker was very helpful in providing the pseudo code and explaining the algorithm of creating mentor groups. I also want to acknowledge Akila Weerapana for assisting with the merging and anonymizing of all the data sets and for being on my thesis committee. On that note, I also want to thank Wendy Wang and Jonathan Tannenhauser for agreeing to be on my thesis committee. I want to give a huge thank you to Casey Pattanayak, my thesis advisor, for all of her support and assistance in this thesis this past year. Thank you for introducing me to the world of Causal Inference, meeting with me every day, and answering all of my questions about the thesis and Statistics. I honestly could not have finished this thesis without you. Lastly, I want to thank all of my friends and family who showed me love and support throughout my time of completing the thesis.

Abstract

This thesis focuses on the effects of a Wellesley student's First Year Mentor's major and First Year advisor's department on the student's choice of major. The Rubin Causal framework was used to draw causal inferences instead of just correlation. Doing so, it is possible to attribute any effect to the type of First Year Mentor or First Year advisor assigned to the student. In the mentoring component, the data was analyzed using Fisher's Test and the Separate Regressions Method. Each method gave some statistically significant results; however they were not practically significant due to their small effect sizes. Our approach to the mentoring component of this thesis illustrates a novel application of rerandomization techniques to a natural experiment. In the advising component, the data was analyzed using subclassification because the assignment of advisors is not randomized. After subclassification, the data was analyzed using a weighted t-test. The results were not statistically significant. Thus, First Mentors and First Year Advisors do not appear to have an effect on a Wellesley student's choice of major.

Table of Contents

1. Introduction

College majors have been a topic of interest for many researchers. Research has ranged from the factors that influence a choice in major (Beggs, et al., 2008) to the factors that encourages a student to major in a science, technology, engineering, and mathematics (STEM) field (Crisp, et al., 2009). Despite the numerous research papers on the choice of majors, there are not many that analyze the effects of faculty advisors and peer mentors on the choice of major. However, there are many studies on the benefits of having peer mentors and faculty advisors, as Campbell and Campbell (1997) found in their work.

In this thesis, I explore the effects of peer mentors and faculty advisors on the choice of STEM versus non-STEM majors of Wellesley College students. In particular, there are two main components in this thesis. The first component focuses on the effect of a student's first year peer mentor's (FYM) major on the student's choice of major. For example, if a student's first year mentor majors in STEM, is the student more likely to major in STEM? The second component focuses on the effect of a student's first year faculty advisor's department on the student's choice of major. For example, if a student's advisor is in a STEM department, is the student more likely to major in STEM?

This thesis is organized as follows. "Background Information on Wellesley College" provides information on Wellesley College. "Literature Review" discusses the previous literature on the choice of major and effects of peer mentors and faculty advisors. The sample and data set will be introduced and explained in the "Sample and Data". "Rubin Causal Framework" and "Background on Statistical Methods for Causal Inference" discusses the statistical theory and methods underlying the thesis that allows for a causal analysis of the two treatments. I then describe the methods of analysis of both components of the thesis in "Randomized Experiment on Peer Mentoring" and "Observational Study on Advising". Within these sections, I discuss the results found. Lastly, I discuss the overall discussions, limitations from the statistical tests, and offer recommendations to the administration based on the results.

2. Background Information on Wellesley College

Wellesley College is a women's college in Wellesley, MA that was founded in 1870 ("Wellesley Facts", 2016). Each year, Wellesley offers many sources of support for students, especially during their first year. In this study, we focus on first year mentors and faculty advisors because every incoming first year is assigned a first year mentor and a faculty advisor. Additionally, each student is expected to meet with their mentors and advisors at least once during orientation week.

First year mentors are "juniors and seniors who provide guidance, support, and information to First - Year Students at Wellesley College" ("First-Year/ Transfer Mentor", 2016). They are also "expected to contribute to new students' personal development as independent thinkers and provide an informed perspective to new students" ("First-Year/ Transfer Mentor", 2016). Mentors must go through an application and interview process before selected as student leaders, and they must attend a mandatory training during the week before orientation. Because of their involvement with first year students, I hypothesize that a student's first year mentor will have an impact on her choice of major.

Each year in August, approximately 50 mentoring groups are randomly formed using an algorithm created by the Library and Technology Services (LTS). Each group has approximately

10-15 first year students and 1 mentor (or 2 for Wellesley Plus, explained below). In general, first year mentor groups are formed based on students' living arrangements. There are three main dorm areas, Tower Court, East Side Dorms, and the Quint, and most groups are formed with students from the same dorm area. The algorithm goes through each dorm, randomizes the list of available mentors, and then goes through the list of students to assign students to mentors, if possible (Ravi Ravishanker, personal communication, March 5, 2017). See section 7 for a detailed explanation of the algorithm. After groups are formed, the balance of the groups are checked, and the algorithm is repeated if the balance is not satisfactory. In addition, there are several specialty mentoring groups that are not based on dorms. The Wellesley Plus group is specifically for students who had a lack of support in high school, and students are selected to apply to the program. They are all placed into one or two groups. Transfer students and Davis Scholars (students that are older than the typical college age) also each have their own mentoring group which is not based on the students' dorms.

Similarly, I hypothesize that a student's first year faculty advisor has an effect on the student's choice of major. Each student is assigned to a faculty advisor in August before orientation week. First year advisors are expected to "help you think most broadly about how to take advantage of the opportunities Wellesley offers" and "help guide her through the process of planning her education" ("First-Year Advising", 2016). In general, faculty advisors meet with their students either individually or in a group during the first week of classes. Afterwards, it is optional for faculty and students to meet throughout the semester, and at times, these first year advisors become major advisors for the student.

Advising pairs are assigned by the Dean of Class of 2017 (currently John O'Keefe) using information provided by students before they enter Wellesley. The assignment process has changed slightly over the years, but in most of the years for which data was available for this thesis, the ultimate goal is to have pairs of student and professor in the same department of interest (John O'Keefe, personal communication, March 17, 2017). Each August, prior to students arriving on campus for Orientation, the Dean assigns students to faculty advisors. To assign these, the Dean first considers the student's intended majors listed on the student's CSS Profile in order to match students with advisors in the same or related departments. The CSS Profile is a form that every student is required to fill out prior to applying for college, and it contains questions regarding the student's financial background and academic interests. For remaining students and students with undeclared intended majors, he uses information from a survey that first year students are asked to fill out during the summer. Wellesley Plus students are assigned to their first year writing and first year seminar professors. In the last entering year included in this thesis (2012), the Dean of Class of 2017 began to assign students to advisors in the same way that Wellesley Plus students are assigned to their advisors. The rest of the students are then assigned based on their intended majors and their responses in their summer form. However, the summer forms were not available for the students in the cohorts studied this thesis, and so certain students were not included (more details in the section 8).

Because this thesis focuses on a student's choice of STEM versus non-STEM major, we have to classify the majors that are considered as a STEM major. To do this, I used Wellesley College's classification of STEM majors in order to have an accurate representation of Wellesley College students. On their "Sciences @ Wellesley" (2016) page, Wellesley lists the following as

STEM majors: Astronomy, Astrophysics, Biological Chemistry, Biological Sciences, Chemistry, Cognitive and Linguistic Sciences, Computer Science, Environmental Studies, Geosciences, Mathematics, Neuroscience, and Physics (2016). In this thesis, I will consider the above majors as STEM majors and every other major will be considered as non-STEM.

It is also important to note two policies that Wellesley has implemented that affects the students' academic interests. In 2003, a grading policy was implemented in order to maintain a class average of a B+ (3.33) in introductory classes with at least ten students ("CCAP Grading Policy FAQ", 2012). This grading policy was implemented in order to maintain a balance of grades between STEM and non-STEM courses as students tend to have lower grades in the STEM courses. In the fall of 2014, Wellesley implemented the Shadow Grading Policy. This policy states that first year students will only receive "pass/ not pass" grades for their first semester classes ("Shadow Grading Policy", 2014). The Shadow Grading Policy was implemented in order to allow first year students to have time to explore and learn about the Wellesley College academic standards without big repercussions. Because these grading policies affect students' grades, I have chosen class years that are similar in terms of exposure to these policies.

3. Literature Review on College Experience

I. Choice of Major

Previous work on the choice of majors ranges from examining the general factors that affect major choices to examining the effect of factors in more specific instances. In many cases, researchers found that there are influences from certain factors with varying degrees of significance and impact.

Beggs, et al. (2008) examined the most important factors in college major choice as reported by students and the importance of each factor. They conducted two studies, one gathering information to determine which factors are important in the choice of major and another which determined the ranking of these factors. In the qualitative study, faculty members conducted interviews with students to gather information about what they thought were important factors in major choice. Beggs, et al. (2008) then analyzed responses from online surveys where students assigned rankings to the factors determined in the interviews. The authors found that the most important factors, in decreasing order, included: student interests, major attributes, job characteristics, financial success, social benefits, and amount of information known by a student (Beggs, et al., 2008).

Trusty (2002) analyzed the effects of external factors on the choice of STEM majors using data from the National Education Longitudinal Study of 1988-1994. He examined the effects of high school courses, race, and gender on choice of major. Trusty (2002) found that women who took math and science classes, specifically calculus, are more likely to pursue a STEM major. However, he found that race did not have a statistically significant effect on choice of major for women. As for males, Trusty (2002) found that being White is a significant factor, and they are only 66% as likely to major in STEM as all other races combined. He also found that men had higher self-efficacy in math and science, but he did not find a positive significance in taking a specific class on the choice of STEM major for males.

Betz and Hackett (1983) studied the choice of a STEM major, but they focused on internal factors such as a student's confidence and self-efficacy in math. Volunteers from undergraduate psychology classes indicated their confidence level in solving math problems, math courses, and math tasks. The authors found that men generally had higher math self-efficacy scores and took more math courses than women. Furthermore, Betz and Hackett (1983) found that students who were more confident in their math skills were more likely to major in STEM majors.

II. Peers in STEM Persistence

Ost looked at the effects of peers, race, gender, and grades on STEM major persistence of a student (2010). Using institutional data from a large research university, Ost looked at the effects of the different factors on persistence in life sciences and physical sciences (2010). Looking at peer effects (peers are defined as those who are in the same classes), Ost used estimated probabilities of persistence of each student to calculate an average by class and found that a for physical sciences, "10% point increase in the propensity of one's peers to persist leads to a 2.05% point increase in the probability of persistence" (2010). Note that Ost does not mean "propensity" in the same sense that I use "propensity" in section 6. He also found that students who were at the bottom quartile of the class were most influenced positively by their peers (3.53% point increase) whereas students in the top quartile are not as affected. However, Ost (2010) did not find that peers were important in life sciences, and he suggests that it is due to the smaller sample size.

III. Faculty Advising

Campbell and Campbell (1997) focused on pairs of faculty advisors and students from a West Coast school in order to examine the positive effects of a faculty advisor. Pairs were formed based on the student's intended major and the advisor's department. More specifically, Campbell and Campbell (1997) analyzed the effects of GPA, retention rates, and graduation rates. To analyze the data and address the lack of randomization, they used matching methods using students' background information to match students who had an advisor to students who did not. The authors then compared GPA, retention rates, and graduation rates between the matched students. Campbell and Campbell (1997) found that faculty advisors have a positive effect on GPA and retention rates. Furthermore, students who met with their advisors more generally had higher GPA and retention rates.

IV. Connections to this Thesis

The previous sections summarize research looking at factors that affect choice of major and the effects of peers and faculty advising. However, from the previous work that I found, there have been none that looked at the effect of faculty advisors and peer mentors on a student's choice of STEM major. In this thesis, I focus on the effect of peer mentors and faculty advisors on a student's choice of either a STEM or non-STEM major. The students of interest are students of Wellesley College, and as previous research found, females are not as likely to major in STEM as males are. Therefore, by focusing on a women's college, I am able to analyze the effects of peer mentors and faculty advisors to determine whether these factors affect females to major in STEM.

4. Sample and Data

This thesis uses Wellesley's institutional data describing students entering in the years of 2006 to 2012. Data was supplied by Wellesley College's Office of Institutional Research (OIR) and the Dean's Office. Data came in electronic and paper form, which I then had to scan, with the help of the Resource Sharing Specialist of the Pforzheimer Learning and Teaching Center (currently Angie Batson), using Optional Character Reader in Adobe. The Faculty Director of the Pforzheimer Learning and Teaching Center (currently Akila Weerapana) then combined and anonymized the all the data. Students within these cohorts have already graduated, so we can analyze the effects of first year mentors and faculty advisors on the majors that the students graduated with. Also, these students have experienced similar academic policies throughout their years at Wellesley; they came after Wellesley's change in grading policy and before Wellesley's Shadow Grading Policy (described in section 2). The data excludes transfer students and Davis Scholars because they are likely to already be on the path toward some major (ie have already taken college courses in the major), which is different from students who came straight to Wellesley after high school. Furthermore, because Wellesley Plus students are assigned mentors and advisors differently than other students, they will be excluded from the analysis. Therefore, the conclusions I draw from this thesis do not generalize to transfer students, Davis Scholars, and Wellesley Plus students.

There are two separate data sets, one for each component of the thesis, because the entering years of students are different. This is because data for specific years was not available. In the peer mentoring component, students came to Wellesley in the years of 2006, 2007, 2009, 2010, 2011, and 2012. In the advising component, students are from the entering years of 2007, 2008, 2009, 2010, 2011, and 2012. The distributions of the students per year are listed in Table 1 and 2. In general, there are approximately 600 incoming students per year and in general, there is an increasing trend in the number of students interested in STEM majors when they fill out their college applications.

Table 1: Distribution of Students in Mentoring Component

	2006	**2007**	**2009**	**2010**	**2011**	**2012**
Number of students	582	587	585	625	571	581
Number of students with STEM at all intended majors	197	195	246	245	224	226
Number of students with non STEM at all intended majors	385	392	339	380	347	355
Number of Distinct Majors	49	51	52	45	48	42
Number of Distinct STEM at all mentors	13	16	14	12	13	17
Number of Distinct non STEM at all mentors	31	30	32	37	34	33

Table 2: Distribution of Students in Advising Component

	2007	**2008**	**2009**	**2010**	**2011**	**2012**
Number of students	586	590	585	625	571	581
Number of students with STEM at all intended majors	195	234	246	245	224	226
Number of students with non STEM at all intended majors	391	356	339	380	347	355
Number of Distinct Majors	51	49	52	45	49	43
Number of Distinct Faculty Advisors in STEM Departments	10	10	11	9	11	8
Number of Distinct Faculty Advisors in non STEM Departments	34	32	38	31	32	30

The variables included in the data set include the student's demographic information, test scores (ACT, SAT, etc.), intended majors at the time students apply to college, and more. For a full list of the available variables, please refer to the Appendix, Table 1. Students at Wellesley College can have up to two majors. Major(s) can either be STEM only, be non STEM only, or in the case of two majors, one be STEM and the other be non STEM. For students who major in both STEM and non STEM, they can list the STEM major first or second when they declare their intended majors. Thus, the mutually exclusive groups of majors are: students who are STEM majors only, students who are non STEM majors only, students who list a STEM major first and non STEM major second, and students who list a non STEM major first and a STEM major second.

Using this information, I created additional variables that provide more information on the student's majors. These variables include: indicator for each major (1 if a student is interested in it, 0 if not), a STEM at all indicator (1 if student lists a STEM major as the first or second interested major, 0 if not), a STEM only indicator (1 if student only lists STEM as interested majors, 0 if not), a non STEM only indicator (1 if student only lists non STEM as interested majors, 0 if not), indicators for listing STEM majors as the first interested major and non STEM as the second interested major, and vice versa, a treatment indicator (1 if mentor or faculty advisor is in STEM at all, 0 if not), and interactions between several variables.

For example, if a student's intended majors are Mathematics or Mathematics and Biology, then she is considered STEM only and STEM at all. If the intended majors are English or English and History, then she is considered non STEM only. If she is interested in Mathematics and English, she is considered as STEM at all and STEM major listed first. If she is interested in English and Mathematics, then she is considered as STEM at all and STEM major listed second.

Our primary outcomes are whether or not the student majors in a STEM major at all, STEM only, non- STEM only, and their cumulative GPA when they graduate.

Next, I will talk about the Rubin Causal Framework and how it applies to this thesis. Afterwards, I will discuss the methods results for each component.

5. Rubin Causal Framework

I. Applied to the Current Context

Causal Inference is the process of determining an effect of a treatment variable on an outcome. This is a crucial concept for this thesis because it allows us to attribute the cause of an effect to the treatment variable. In the context of this thesis, causal inference allows us to say whether a student is more likely to major in STEM because her first year mentor is a STEM major. Without causal inference, it is only possible to say that two factors are correlated to one another, and correlation does not imply causation. To draw causal inferences from this study, we use the Rubin Causal Framework.

The Rubin Causal Framework defines a causal question in terms of units, treatments, covariates, and potential outcomes. A "unit" is an object, person, collection of objects, etc. in which an action is applied to at a particular point in time (Rubin, 1974). In this thesis, the unit is a Wellesley College student. A treatment is an action that is applied to a unit (Imbens and Rubin, 2015). In the mentoring component, the active treatment is that a student is assigned to a mentor who majors in STEM at all, and the control treatment is that a student is assigned to a mentor who does not major in STEM at all. In the advising component, the active treatment is that a student is assigned to a STEM faculty advisor and the control treatment is that a student is assigned to a non STEM faculty advisor. Covariates are variables reflecting background information that is collected before the treatments are assigned and thus cannot be affected by the treatments assigned (Rubin, 1974). In this thesis, covariates consist of all the information that is collected before the first year mentors and first year advisors are assigned to students in the summer. Some covariates include intended majors, SAT scores, and ACT scores.

Rubin (1974) defines a "potential outcome" as the outcome for a unit at a particular time after a particular treatment is applied. The number of potential outcomes for each unit is the same as the number of treatment options. For example, each student in the mentoring component has two potential outcomes: her eventual major if she is assigned a STEM mentor and her eventual outcome if she is assigned a non STEM mentor. Whether or not a student majors in a STEM major given that they had a STEM mentor is a potential outcome under the active treatment and is a potential outcome under the control treatment if the student did not have a STEM mentor. In the advising component, the two potential outcomes are: a student's eventual major if she is assigned a STEM advisor and her eventual outcome if she is assigned a non STEM advisor.

The causal effect is defined as a comparison between the potential outcomes under each treatment (Rubin, 1974). The "estimand" is something that can be calculated from the data. In this thesis, I use the mean of difference in potential outcomes for each unit as the estimand, which is a typical definition of causal effect. Thus, if we observed both potential outcomes for each unit, then, to determine the causal effect, I calculate the difference between the mean potential outcomes of the active treatment and the mean potential outcomes of the control treatment.

II. The Science Table

The visual framework for the Rubin Causal Framework is the Science Table (example shown in Tables 3a and 3b), which is a table that represents the situation-- each row represents a unit and each column represents covariates and potential outcomes (Rubin, 1974). The Science Table ensures that the problem can be framed in a way that is appropriate for causal inference.

Table 3a: Example Science Table

Unit	***Covariate***	***Potential Outcome under Control Treatment***	***Potential Outcome under Active Treatment***
Unit	**Intended Major**	**Y(non-STEM)**	**Y(STEM)**
1	Math	STEM	STEM
2	Math	non-STEM	STEM
3	English	non-STEM	non-STEM
4	English	STEM	non-STEM
5	Biology	non-STEM	STEM
6	Biology	STEM	non-STEM

Table 3a shows an example of a Science Table looking at the effect of a first year mentor's major. The unit IDs are listed in the first column, and the unit's intended major, a covariate, is listed in the second column. The next two columns show the potential majors of each unit, given that the unit was assigned to a mentor who is a non STEM major and given that the unit was assigned to a mentor who is a STEM major, respectively.

Table 3b: Example Science Table with Binary Indicator

Unit	Intended Major	Y(non-STEM)	Y(STEM)	Causal Effect
1	Math	1	1	0
2	Math	0	1	1
3	English	0	0	0
4	English	1	0	-1
5	Biology	0	1	1
6	Biology	1	0	-1

In Table 3b, the first two columns are the same as in Table 3a. In the potential outcomes of active and control treatment, I substituted 1 for a STEM major and 0 for a non-STEM major. Doing so, I can calculate the causal effect by calculating the difference, as shown in the Causal Effect column. A positive effect means that a STEM mentor causes a student to major in STEM, a negative effect means that a STEM mentor causes a mentor to major in non STEM, and a zero effect means that a STEM mentor has no effect on the student's choice of major.

III. SUTVA

If we are to assume that the Science Table is a sufficient representation of the situation, the Science Table must satisfy the Stable Unit Treatment Value Assumptions (SUTVA). There are two parts of SUTVA. Part I of SUTVA states that treatments are clearly defined and there are no different versions of treatment for each unit (Imbens and Rubin, 2015). This allows for treatments and potential outcomes to be represented by columns in the Science Table. If SUTVA I does not hold, the two columns for potential outcomes is not enough. More columns would need to be added. For example, if we looked at three types of peer mentor majors: STEM, Social

Sciences, and Humanities, three columns would be needed in the Science Table. However, the causal effect as a comparison between three columns is harder to calculate, and method of analysis become much more complicated.

In the first year mentor component, I have defined the treatments keeping in mind of double majors and excluding minors. As explained earlier, active treatment is when the student is assigned a mentor who is a STEM major at all and control treatment is when the student does not have a mentor who is a STEM major at all. Using this definition of active and control treatment, SUTVA I holds because those are the only treatments that a student can receive. In the advising component, the active treatment is assigning a STEM faculty advisor to a student and the control treatment is assigning a non-STEM faculty advisor to a student. Because the data only contains one department for each advisor, there is no variation to the treatments. Thus SUTVA I is valid in both components

Part II of SUTVA states that the potential outcomes of a unit do not depend on the treatment of another unit (Imbens and Rubin, 2015). If SUTVA II does not hold, then the Science Table is not sufficient for the situation because additional columns would be needed for the potential outcomes of the interactions between the units. For example, if whether or not a student majors in STEM depends on who her friend's mentor is, then columns for four potential outcomes are needed for this student: (1) whether she is a STEM major if she and her friend both have STEM mentors, (2) whether she is a STEM major if she has a STEM mentor but her friend does not, (3) whether she is a STEM major if she does not have a STEM mentor but her friend does, and (4) whether she is a STEM major if she and her friend both do not have STEM mentors.

It is harder to say that SUTVA II is true, but in this thesis, I assume that SUTVA II holds. In both components, students interact with their friends and other students who may have an influence on what the students major in. Because of this, whether or not the friend has a STEM mentor and whether or not the friend has a STEM advisor might affect what the student eventually majors in. However, if SUTVA II does not hold, a network of the student's friends is required in order to properly analyze the results. Instead, to make the analysis simpler, I assume that SUTVA II holds. By assuming SUTVA II holds, I assume that whether a student's friend has a STEM mentor or advisor does not affect what the student majors in. The argument underlying this assumption is that a student's choice in major is not easily influenced by their friends. Thus, the student's friend's mentor and advisor do not affect what the student majors in.

IV. Fundamental Problem of Causal Inference

Recall that in the Science Table (Table 3), there are columns for potential outcomes of each treatment. However, in reality we could never observe this Science Table because in one moment in time, a unit cannot be assigned both a control and an active treatment (Rubin, 1974). This is the "Fundamental Problem of Causal Inference" (Holland, 1986). Because of this problem, it is only possible to estimate the causal effects; we can never calculate them in the way that we could when looking at Table 3 (Imbens and Rubin, 2015). To estimate the causal effects, the missing potential outcomes need to be imputed so that a difference in potential outcomes can be calculated. Revisiting the example in Table 3, students can only be assigned a STEM at all mentor or a non STEM at all mentor. Thus, in reality, the potential outcomes of the treatment that they did not receive are not observed. Then, I cross off the unobserved potential outcomes.

Using the data that is observed, we can attempt to estimate and impute these missing values of potential outcomes.

Table 4: Fundamental Problem of Causal Inference

Unit	**Intended Major**	**Assignment Vector**	**Y(non-STEM)**	**Y(STEM)**	**Causal Effect**
1	Math	STEM	?	1	?
2	Math	non-STEM	0	?	?
3	English	STEM	?	0	?
4	English	non-STEM	1	?	?
5	Biology	STEM	?	1	?
6	Biology	non-STEM	1	?	?

V. Assignment Mechanism

Because of the Fundamental Problem of Causal Inference, which potential outcome we observe for each unit depends on how treatment is assigned. The way in which treatment is assigned is the "assignment mechanism" (Imbens and Rubin, 2015) It is important to understand the assignment mechanism in order to impute missing potential outcomes and draw inferences about the Science Table.

a. Randomized Experiment

To be classified as a randomized experiment, the assignment mechanism must be probabilistic and known. An assignment mechanism is probabilistic if every unit has a probability, p_i, of receiving the active treatment, where p_i is between 0 and 1, but not equal to 0 or 1 for any unit (Imbens and Rubin, 2015). An assignment mechanism is known if the person analyzing the data knows how the treatment was assigned to units.

b. Observational Studies

For an observational study, the assignment mechanism is not known. However, observational studies can also have the properties described in part c.

c. Other properties

There are other properties that can describe an assignment mechanism. These are the unconfoundedness property and the individualistic property. An assignment mechanism is unconfounded if it does not depend on the potential outcomes, given the observed covariates, ie the covariates are the only information needed to determine the assignment (Imbens and Rubin, 2015). If unconfoundedness does not hold, there is a correlation between the assignment mechanism and the potential outcome, even when the observed covariates are taken into account. For example, a randomized experiment in which everyone is assigned a treatment at the same time is unconfounded. However, an observational study in which we do not observe the covariates that determine the treatment assignment is confounded because the units who get active treatment are likely to have a different set of potential outcomes than the units who got control treatment. A "unit level probability" is the probability that a unit is assigned an active treatment. An assignment mechanism is individualistic if it meets two conditions. The first part says that each unit's probability of assignment to active treatment is a function of only that unit's row in the Science Table. The second part states that the probability of an assignment vector is proportional to the product of the unit level probabilities (Imbens and Rubin, 2015). An assignment vector is a vector of treatment assignments for all units.

d. Assignment Mechanism in the thesis

Clearly, the assignment of mentors is randomized because the mentoring groups are created using a randomized algorithm created by the Library and Technology Services office (full details in Section 7). First, we know what the process of treatment assignment is. Second, since the algorithm randomizes a list of mentors when creating groups, the assignment mechanism is probabilistic. This is because conditional on the housing assignments, every student has a non-zero and non-100% chance of either being assigned to a STEM mentor or non-STEM mentor. Since the assignment is randomized all at once, the only information needed to make the treatment assignments is the dorm information of the students. This information is available, and so the mentoring component is unconfounded. However, there is reason to worry that the mentoring assignment mechanism is not individualistic. The first part of individualisticness holds, conditional on the dorm area of a student, because the probability that a student is given the active treatment depends on the number of mentors assigned in the dorm area and not the other students. It is less clear whether the second part of individualisticness holds because the algorithm for assigning mentors is complex. We cannot say for sure that the peer mentoring component is a Classical Randomized Experiment. This is one reason that we focus on randomization- based and multiple imputation based inference, as explained in Section 7, rather than relying on t-tests and large sample theory.

The advising component of the thesis is an observational study because faculty advisors as assigned via a subjective process that has not been written down as an algorithm or equation. The assignment mechanism is unknown. It is also not probabilistic because students with certain intended majors have a 0 or 1 probability of being assigned a particular type of advisor (eg. a student who have intended majors in Physics and Mathematics are essentially never assigned non STEM advisors). The advising assignment mechanism is also not unconfounded given the covariates available to us because for leftover students and students who have undeclared intended majors, the Dean of Class of 2017 used information on the form that the student filled out during the summer. However, these forms were not available to me, and so there might be a correlation between the treatment assigned and the potential outcome because I cannot take into account the information provided in the forms. Even if a comparison is made on people with undeclared majors who are similar on all of the covariates that I did observe, there is still the likelihood that those who got STEM advisors mentioned STEM on their forms and those who did not get STEM advisors did not mention STEM on their unobserved forms. The confounded assignment mechanism implies that conditioning only on the observed covariates will not be sufficient for drawing causal inferences. To partially account for this problem, I removed the students who had undeclared majors first and second choices. As for the other students whose advisors were assigned using the form, I assume that unconfoundedness holds. Similar to the mentoring component, it is difficult to determine whether the assignment mechanism is individualistic. Because this component is an observational study, the underlying unit level probabilities are not known.

6. Background on Statistical Methods for Causal Inference

In this thesis, the two components consist of a randomized experiment and an observational study. In this section, I further discuss statistical methods appropriate for each type of study.

I. Randomized Experiments

a. Causal Inference in Randomized Experiments

Conditional on the covariates related to treatment assignment probabilities, treatments are assigned to the units randomly in a randomized experiment. Thus, distribution of observable and unobservable covariates, such as race, test scores, and interests are expected to be the same in the two treatment groups. The only difference between the units from the two treatment groups is the treatment they receive. Because this is the only difference, an unbiased estimate of the causal effect can be estimated. If an effect is observed, I can attribute the cause to the treatment received. Thus, it is straightforward to draw causal inference of the active treatment versus the control treatment in a randomized experiment.

b. Fisher Randomization Test

The Fisher Randomization Test is a nonparametric test that does not require any assumptions besides that the assignment mechanism is known (Fisher, 1925, 1935). The test looks at how likely the observed difference in potential outcomes from the data is. To do so, we compare it to a distribution of differences, under the sharp null hypothesis, which is calculated by looking at all the different possible assignment vectors. A "sharp null hypothesis" is a null hypothesis that specifies a value for each missing potential outcome, and so we can fill out missing potential outcomes using the sharp null hypothesis. For example, under the sharp null that having a STEM mentor has a zero effect on what any particular student majors in, then each student's missing potential outcome is equal to her observed potential outcome for the opposite treatment. Because we look at all of the possible treatments assigned to all units, the assignment mechanism must be known in order to conduct a Fisher Test. The possible assignment vectors can be computed mathematically, but it is more common to conduct the test by simulation, randomly creating possible assignment vectors according to the assignment mechanism.

c. Fisher Interval

The Fisher Interval is the set of possible values for the unit-level causal effect assumed by the null hypothesis that would not be rejected using a randomization test, assuming that the unit-level causal effect is the same for each unit. To find the values that are included in the interval, the randomization test is conducted repeatedly for varying sharp null hypotheses (Imbens and Rubin, 2015). For any particular hypothesized unit-level causal effect (e.g. the difference between the potential outcomes is 3 for each unit), the missing potential outcome values are filled in (e.g. we hypothesize that a person whose potential outcome under control was 1 would have a potential outcome, under the active treatment, of 1+3 = 4). Then, the randomization test is conducted with this null Science Table, and a p-value is calculated. If the sharp null hypothesis can be rejected, the hypothesized unit-level treatment effect is not included in the interval.

d. Rerandomization

Rerandomization is a recently proposed study design that attempts to reduce the impact of chance imbalances in randomized experiments (Morgan and Rubin, 2012). Although covariates are expected to be balanced between active treatment and control groups, often times this is not the case. Given that criteria was determined in advanced, rerandomization is "a tool that allows us to draw from predefined set of acceptable randomizations" (Morgan and Rubin, 2012). The main idea is to randomize treatments to units repeatedly until the resulting covariate balance between the active treatment and control groups meets preset criteria. In other words, even if treatment assignments are randomized and rerandomized multiple times, only a smaller set of acceptable treatment assignments are possible. Furthermore, randomized based methods, such as the Fisher Test, can be used in rerandomized experiments: instead of examining the distribution of the difference in means under all randomizations that could be obtained, the set of randomizations used for the Fisher Test is restricted to randomizations that would have met the predefined acceptability criteria (Morgan and Rubin, 2012).

e. Separate Regressions Method

The Separate Regressions Method is a Bayesian method that explicitly and repeatedly imputes the missing potential outcomes using a pair of regression models. This method is similar to the method that Belson proposed (1956). The Separate Regressions Method assumes that the regressions model used is true, which is an assumption that might not be true. However, I assume that the model holds in order to proceed to take into account the relationship between covariates and outcomes and to add precision. Covariates are chosen as predictors in order to generate a model with the best fit to predict potential outcomes. Crucially, the method involves fitting separate models to predict the potential outcomes under active treatment and the potential outcomes under control, given the observed covariates. We do not assume that the potential outcomes under each treatment are related to covariates the same way, although we do assume independence between the potential outcomes. This method differs importantly from running a regression of observed outcomes on the covariates.

Typically, a regression model using covariate information from units who received active treatment is used to impute missing values for units who received control treatment, and vice versa. The data set is first split into units who received active treatment and units who received control. Then, for each potential outcome, we fit two models, one using the units who received active treatment and one using the units who received control treatment. Using these models, we impute missing potential outcomes by using the model run on the units that received opposite treatment but the covariates from the units we are trying to impute potential outcomes for. The difference in means in potential outcomes is then calculated to estimate the causal effect. This process is then repeated, say, 1000 times in order to generate a distribution of values for the difference in means. A 95% Bayesian interval is then obtained based on the 2.5th and 97.5th percentiles of the distribution.

II. Observational Studies

As mentioned before, it is straightforward to draw causal inferences from a randomized experiment because the only difference between two treatment groups is the type of treatment received, conditional on any covariates used for the randomization. However, it is still possible to draw causal inference in an observational study if we observe the appropriate covariates. To do so, the observational study must be designed to parallel a randomized experiment. First, we have to assess whether we have collected all of the covariates likely related to the treatment assignment and the potential outcomes. To determine this, we consult the people who make the treatment decisions. Secondly, the distribution of the observed covariates in the two treatment groups should be very similar so that the only difference between the units is the treatment assigned.

a. Subclassification (and Matching)

To design the observational study so that a balance of covariates is achieved, subclassification and matching methods are commonly used. Subclassification is the stratification of similar units into groups based on their covariates. Matching is the process of pairing units from different treatment groups based on covariates. The purpose of subclassification and matching is to ensure that the units of comparison in the two treatment groups differ only by the type of treatment they receive. Since the only difference is the type of treatment that units receive, any differences can be attributed to the type of treatment, assuming that all of the necessary covariates were observed and included.

b. Creating Covariate Balance

To achieve a balance of covariates, we have to use subclassification or matching methods. For example, say we wanted to balance the SAT Math scores within the two treatment groups. For subclassification, subclasses can be created by grouping students with scores below the median SAT Math scores and grouping students with scores above the median SAT Math score. Then, within each subclass, the mean SAT Math score should be balanced between units that received active treatment and units that received control treatment. For matching, students in one treatment group are paired with another student in the other treatment group who has the same SAT Math score. Then, within pairs, the mean SAT Math score is balanced between units that received active treatment and units that received control treatment. However, it is hard to balance multiple covariates at the same time. Instead, propensity score estimation is often used for subclassification or matching.

A “propensity score” is the probability that the unit is assigned the active treatment, conditional on covariates. To estimate the propensity scores, we can use a logistic regression where the dependent variable is the actual treatment assignment of each unit. The independent variables used in this logistic regression are covariates that we wish to prioritize. The fitted probabilities from this model can be used as estimated propensity scores, and each unit will have an estimated propensity score, just as it has a values for every other covariate. Subclassification or matching on estimated propensity scores can establish balance on many covariates simultaneously.

c. Analysis of outcomes

Once the covariates between the units of active and control treatment are balanced, the data can be analyzed as if it came from a randomized experiment. In subclassification, units of each treatment type are compared within each subclass. In matching, units are compared in pairs. In both subclassification and matching, the comparison of potential outcomes is done on units with similar covariates, and so the outcome can be attributed to the active treatment. To analyze the data, a t-test or an appropriate extension of a t-test can be used to compare the outcomes of the two treatment groups.

7. Randomized Experiment on Peer Mentoring

I. Methods

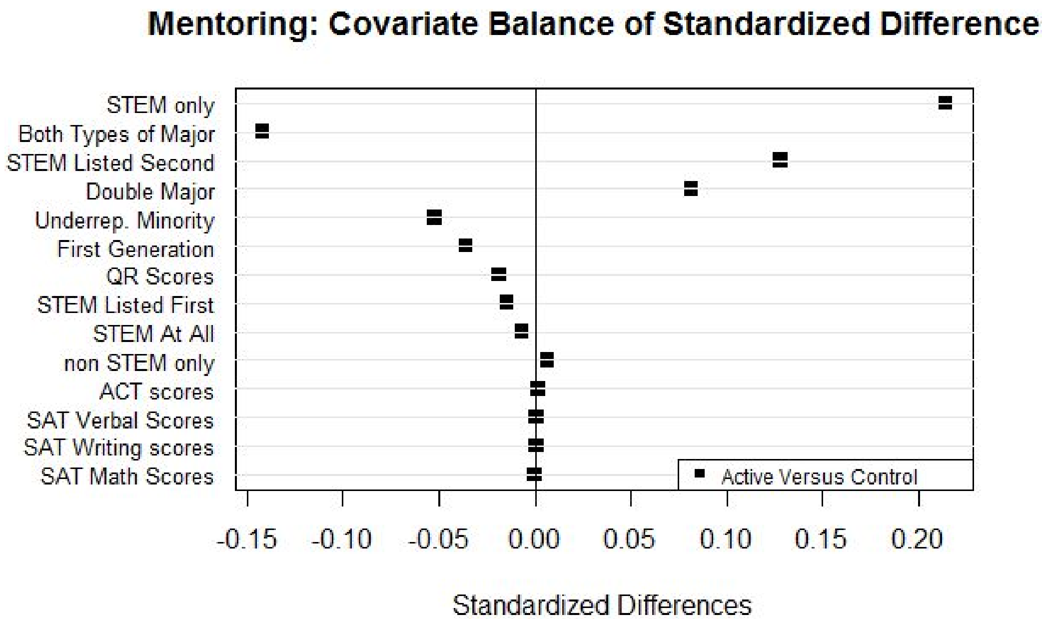
Note that during the design stage of the thesis, it is common statistical practice to the temporarily remove the outcomes from the data set (Rubin, 2007).

a. Covariate Balance

As mentioned before in Section 6, if treatment assignments are randomized, it is expected that the covariates of both treatment groups are balanced. To ensure that this is true, I created Love Plots to graphically show the difference in covariates between the two groups. Note that the plots depict the standardized differences. To calculate the standardized differences, a normalizing coefficient of $\sqrt{\frac{s_T^2 + s_C^2}{2}}$ is used to multiply the difference in means of the variable (as used by Austin (2009)). The s_T^2 is the estimated variance of the covariate for the treatment group and s_C^2 is the estimated variance of the covariate for the control group.

Figure 1 shows the Love Plot of the covariate balance for the mentoring component. The covariates shown are the prioritized covariates. For these covariates, achieving excellent covariate balance is a priority. In general, this seems to be true. The covariates with the largest differences are STEM only indicator and both types of major indicator (student is interested in two majors, one STEM and one not). However, their differences still only have a magnitude of approximately 0.2. Because of the approximate balance of covariates, I can proceed with the analysis of randomized experiments by calculating the difference in mean potential outcomes.

Figure 1: Love Plot of Covariate Balance of Standardized Differences for Mentoring Component



First, a randomization-based Fisher test that only assumes that we know the assignment mechanism was used to analyze the effect of first year mentors. Then, the Separate Regressions Method was used in an attempt to adjust for chance imbalances in the Fisher Test.

b. Random Algorithm Used to Form Mentoring Groups

To analyze the data for the mentoring component of this thesis, I used a Fisher Randomization test. To obtain a set of possible treatment assignment vectors, I replicated the algorithm that is used currently by the Library and Technology Services to form mentoring groups in the summer. However the algorithm that I used is not an exact replicate because the actual code for the algorithm was not made available to me. Instead, through email correspondence with Ravi Ravishanker (March 5, 2017), he sent a pseudo code of the algorithm. Using the pseudo code, I coded the following algorithm.

The algorithm first looks at where the mentor lives and then randomly chooses n dorms out of a list of all possible dorms that the mentor can be assigned. The dorms that a mentor can be assigned depends on where she lives. If she lives in Tower Court, she can have mentees from Severance, Tower Court East, Tower Court West, and Claflin. If she lives in the Quint, she can have mentees from Pomeroy, Shafer, Cazenove, Beebe, and Munger. If she lives in the East Side, she can have mentees from McAfee, Bates, Freeman, Stone, Davis, and Dower. In this thesis, I chose n to be four because the minimum number of available dorms is four and the maximum is six. The four dorms that are assigned to a mentor are dorms where mentors can have mentees from. The capacity of students from each dorm that a mentor can have is then calculated by dividing the number of students in that dorm by the number of mentors who are assigned to it.

Afterwards, the list of students, separated into entering class years, is ordered by dorm and room number. The algorithm then goes through the list of dorms and mentors, and if the dorm is available to the mentor, we go through the list of students. If the student's dorm is available to mentor and the mentor's capacity for that dorm has not been filled yet, the student is placed inside the mentor's group and removed from the student list. This goes until all mentor's capacities of their n assigned forms are filled. Then, if there are remaining students to be placed, an additional dorm from their original available list is added into their assigned list. Using the updated available dorms list, the algorithm goes through the mentors and students again to place as many students into groups as possible. Note that because this algorithm is not an exact replication of the actual algorithm, there are some students who are not placed into groups.

c. Rerandomization in Mentoring Component

In practice, after the groups were created, the balance of students (balance on dorms, balance on number of students, etc.) is checked by the people who form the mentoring groups. If the balance is not satisfactory, the random algorithm is used again to form new groups until the balance of the groups is achieved. Therefore, the mentoring component of this thesis is not only a naturally randomized experiment, but a naturally rerandomized experiment, and we analyzed it accordingly. We know of no other naturally occurring rerandomized experiments that were analyzed according to the procedures developed for experiments prospectively designed by rerandomization.

To account for the naturally occurring rerandomization, I set criteria before analyzing the balance of the treatment groups in order to determine which treatment assignments are acceptable and which are not. The criteria that I set ensured that 1) the size of mentoring groups are approximately the same and 2) the number of students that are not placed into mentoring groups are minimized. These criteria are intended mimic the way that Library and Technology Services chooses an acceptable mentor group assignment each year. Through discussion with the Interim Dean of Students (currently Lori Tenser), randomizations are repeated approximately three to four times (Lori Tenser, personal communication, October 17, 2016). Because I simulate 4000 simulations, I adjusted the criteria needed per year, with the goal of having at least 15% of the simulations be acceptable. Table 5 shows the criteria that were set per class year. Four thousand treatment vectors were simulated, and the number of acceptable treatment vectors varied, as shown in the table.

Table 5: Criteria and Number of Acceptable Simulations per Year

	2006	2007	2009	2010	2011	2012
Difference in maximum and minimum sized groups	9	11	13	13	13	13
Maximum number of students left who were not placed in a group	17	65	52	52	35	42
Number of accepted simulated treatments	661	563	1008	887	594	1664

The randomizations were carried out separately by year, but our data set includes the students from all years, and so we concatenated the simulated assignment vector to obtain assignment vectors with length equal to the total sample size. To accomplish that, we needed to have the same number of acceptable treatment assignment vectors each year. I chose the minimum number of simulations kept (563) as the number of simulations of treatment assignments to be used for all the entering class years. Using the 563 simulated treatment assignments for each entering class year, I then calculated the difference in potential outcomes to generate a distribution of causal effects. Recall that the outcomes of interest are whether the student majored in a STEM major at all, STEM only, non STEM only, and the cumulative GPA. Lastly, I compared the observed difference in potential outcomes to the distributions to obtain a p-value to determine whether the null hypothesis of zero effect can be rejected or not.

Note that Fisher Intervals cannot be generated in this mentoring component. This is because the only possible values that could be imputed are 0 and 1. Thus, it does not make sense to test for causal effects, such as 3, because imputing values to have a causal effect of 3 does not make sense in this context. Thus, there are only limited values that could be tested for the Fisher Interval. Davis Watson has begun to explore the idea of Fisher Intervals for binary values (slides on file with me, 2012). However, because the concept is still relatively new, we decided not to explore this idea. Instead, we use the Separate Regressions Method to generate Bayesian intervals.

d. Separate Regression in Mentoring Component

In order to use covariates to adjust for chance imbalances on prioritized covariates, I used the Separate Regressions method as another analysis on the mentoring component. To analyze the data using a Separate Regressions method, I first separated the data into units who received active treatment and those who received control treatment. Then, I did trial and error and ran several models with different predictors in a logistic regression model on each of the potential outcomes of interest. After approximately ten models were run for a treatment group for a potential outcome, I chose the model with the minimum AIC (Akaike's An Information Criterion) in R. The AIC determines whether the model is a good fit of the data or not. A lower AIC signifies a better fit. Table 6 shows the predictors and their coefficients for each model. Again, note that I separately modeled the missing potential outcomes under active treatment and the missing potential outcomes under control treatment for each potential outcome variable.

Using these regression models, I simulated 1000 causal effect estimates by imputing the missing potential outcomes. For the units that received active treatment, I used the logistic models predicting potential outcome under control treatment that were derived using the units who actually received control treatment to calculate a fitted probability that each of these units would have a potential outcome under control equal to 1. Then, the missing potential outcomes under control were imputed based on these probabilities by simulating independent Bernoulli draws. The reverse was done to impute the missing potential outcomes for units who received control treatment using models derived from units under active treatment. After the missing potential outcomes are imputed, the differences in mean potential outcomes in active treatment versus control treatment are calculated. These steps were then repeated 1000 times in order to obtain a posterior distribution of differences in means. As for the cumulative GPA, a linear regression was used to impute the missing potential outcomes, and then random noise, with residual variance estimated by the model, was added to generate a distribution of 1000 estimates of the difference in mean potential outcomes

Once the simulations were generated, the observed difference of each potential outcome was compared to the respective distribution. A p-value was calculated by looking at the proportion of estimated causal effects that are greater than the observed difference. Lastly, Bayesian Credible intervals were generated by using the 2.5th percentile and the 97.5th percentile values of differences in mean potential outcomes as the upper and lower bounds of the interval.

Table 6: Predictors and Coefficients for Separate Regression Models

	STEM At All	**STEM At All**	**STEM only**	**STEM only**	**Non STEM only**	**Non STEM only**	**Cumulative GPA**	**Cumulative GPA**
Treatment	STEM	Non STEM	STEM	Non STEM	STEM	Non STEM	STEM	Non STEM
First Year Cohort		.009***	.054	.044	-.051	-.076*	.008	.008*
White	-.172							
Asian	.151		.335					-.007***
Underrepresented Minority	-.387		888				-.138***	-.071**
Pre Med At All Indicator	1.049	1.088***	1.129	1.122***	-.919***	-1.026***	-.083*	-.046
International Relations At All Indicator	-.361***							
Non STEM only	-1.284	-.319	-.557	-.664*	1.265***	.333		
QR Score and SAT Math interaction	.0002**	-.0001	-.0001					
STEM major 1	.878**	1.72***	1.328***	1.261***	-.795**	-1.673***		-.028
STEM major 2		1.077***	1.015***	.876***		-1.036***		-.016
SAT math	-.0006	.555**	-.0007		-.006**	-.003**		
vote_total		-.019**						
foreign		-.189						
QR Score		.227**	-.011	.161***	-.089*	-.159***	.027***	.026***
SAT writing							.0009***	.0007***
SAT verbal							.0004*	.0004***
Math At All Indicator					.372			-.079**
Black								-.151***
Latina								-.105**
ACT					-.004			.009*
Biology At All Indicator					-.405			
English At All indicator						.278		
Classics At All indicator						.541		
First Generation						.003		

(Values are significant at: *** .001, ** .01, * .05

II. Results

The distribution of type of majors that students majored in by the type of majors they intended to major in is shown in Table 7. The intended majors are shown in the columns and the outcome majors are shown in the rows. The first sub table for STEM At All intended major and STEM At All outcome major contains the keys of the rows and columns. The treatment is along the columns of the sub tables and the outcomes are along the rows of the sub table. Overall, it seems that even though the students had a general idea of what they wanted to major in, not everyone followed through. For example, looking at the STEM only intended major column and STEM only outcome major row, there are 177 students across 6 years who ended up not majoring in STEM only majors. Furthermore, of the people who were STEM only intended majors and had a non STEM mentor, 45% of the students ended up not majoring in STEM only majors.

Table 7: The Distribution of Major Types by Treatment and Treatment Types of Intended Majors

Intended Majors → Outcome Majors		STEM At All		STEM only		Non STEM only		STEM major listed first		STEM major listed second	
STEM At All		Treatment - 0	1								
	0	0	0	78	50	1176	490	160	50	126	47
	1	657	292	185	90	190	90	194	72	88	40
STEM only		164	72	119	58	1238	521	207	74	140	56
		493	220	144	82	128	59	147	48	74	31
Non STEM only		657	292	185	90	190	90	194	72	88	40
		0	0	78	50	1176	490	160	50	126	47
STEM major listed first		75	31	93	53	1210	506	179	58	133	51
		582	261	170	87	156	74	175	64	81	36
STEM major listed second		89	41	39	10	403	166	54	25	36	21
		95	46	22	12	38	18	27	11	8	5

a. Fisher Randomization Test

Table 8 contains the actual differences in means as well as the right sided p-values from the randomization tests in parentheses. A right sided p-value is the probability that, given the null hypothesis, we see data at least as great as the actual observed difference in mean potential outcomes.

Table 8: Actual Differences in Means and P-values for Fisher Test for Mentoring Component

	All Years	2006	2007	2009	2010	2011	2012
Number of Students	3424	581	563	561	601	549	569
STEM At all Indicator	.015 (.247)	.021 (.401)	.005 (.638)	-.011 (.579)	.061 (.085)	.033 (.144)	-.015 (.679)
STEM only Indicator	.012 (.309)	.015 (.362)	.005 (.608)	-.029 (.815)	.049 (.062)	.037 (.135)	-.005 (.551)
Non STEM only indicator	-.015 (.753)	-.021 (.401)	-.005 (.396)	.011 (.42)	-.061 (.915)	-.033 (.857)	.015 (.322)
Cumulative GPA	-.011 (.861)	.015 (.442)	-.007 (.641)	-.006 (.487)	-.077 (.977)*	-.035 (.858)	.029 (.199)

Note: * means that it is statistically significant at the two sided cutoff at the.05 level

Looking at the observed difference in means, all of them are approximately zero. The maximum difference in magnitude between the two treatment groups was the difference in cumulative GPA for students who came to Wellesley in 2010. The difference suggests that in the data, a student with a STEM mentor, the student's GPA drops on average by .077 points. In fact, this is also the only statistically significant difference at the two-sided cut off, with a right sided p-value of 0.977. The students, in the data, entering Wellesley in 2007 seemed to be least affected by their first year mentors' majors because magnitude-wise, they have the smallest differences in means between the two treatment groups.

b. Separate Regressions Method

The results from the Separate Regressions Method are shown in Table 9. The estimated difference in means and Bayesian Credible Intervals are shown for each year in parentheses.

Table 9: Estimated Difference in Means with Bayesian Credible Intervals

	All Years	2006	2007	2009	2010	2011	2012
Number of Students	3424	581	563	561	601	549	569
STEM At all Indicator	.012 (.0008, .0242)*	.022 (-.0086, .0534)	.016 (-.0125, .0479)	.017 (-.0143, .0463)	.017 (-.015, .0466)	.014 (-.0182, .0437)	-.017 (-.0485, .0123)
STEM only Indicator	-.002 (-.0141, .0107)	.005 (-.0247, .0339)	-.004 (-.0328, .0259)	-.007 (-.0359, .0223)	.007 (-.0216, .0353)	.002 (-.0265, .0331)	-.012 (-.0406, .0156)
Non STEM only indicator	-.013 (-.0251, -.0003)*	-.008 (-.0361, .0207)	-.009 (-.0409, .0195)	-.017 (-.0481, .0143)	-.022 (-.0516, .0083)	-.021 (-.0528, .0091)	.002 (-.0264, .0334)
Cumulative GPA	-.011 (-.0491, .0236)	-.003 (-.084, .0736)	-.007 (-.086, .0685)	-.007 (-.0879, .0824)	-.029 (-.1113, .0479)	-.006 (-.0897, .0757)	.002 (-.0934, .0717)

Note: * means the interval does not include zero

Looking at the intervals, most of them include zero. This suggests that the data is consistent with the null hypothesis of zero effect. The only intervals that do not include zero are the intervals for STEM At all indicator and for Non STEM indicator using the regression model ran on all subclasses (.0008, .0242) and (-.0251, -.0003), respectively. However, even though they does not include zero, the lower and upper bound, respectively, are very close to zero. In general, the intervals are fairly similar in magnitude and in width. Even though there were statistically significant effects, the magnitudes of the intervals and estimated differences are so small that they are negligible. Furthermore, because of the Multiple Comparisons Problem, even though two intervals did not include zero, by chance, this could occur, and so it is possible that the results are not statistically significant.

8. Observational Study on Advising

I. Methods

Similar to the mentoring component, the outcomes were removed from the data set in order to prevent any biases as it is common statistical practice (Rubin, 2007). In particular, someone other than me, my thesis advisor, removed the outcomes from the data set before I opened the data set. I only accessed the outcomes after finalizing the observational study design. Because the outcomes are removed, I am able to explore the covariate data in any way I want without having to worry about biasing the result.

a. Discarding units

As mentioned before, the Dean of Class of 2017 uses information found on the summer form that students complete to assign advisors to students who had undeclared intended majors. Because of this, students who are undeclared are omitted from the advising component of the thesis. Furthermore, there is only a small number of Wellesley Plus students, and so it is not possible to compare the potential outcomes of these students as there is a great imbalance in the number of students who received active and control treatment. Additionally, we found that it was not possible to draw causal inferences about the impact of advising on students who have intended majors that are non STEM only or listed a STEM major as a second intended major. These students are omitted from the advising component. This is because it is very unlikely for a student who listed non STEM intended majors to be assigned a STEM advisor. When a non STEM student is assigned a STEM advisor, it is because the student is a Pre- Med student (for this thesis, Pre- Med is considered a non STEM major because it is not one of the majors classified as STEM by Wellesley as mentioned in Section 2). Thus, there is a great imbalance in the proportion of Pre- Med students in both treatment groups. This is also true for students who listed an intended STEM major second. Thus, the advising component only focuses on students who declared only STEM intended majors and those who declared a STEM major first. This means that the advising results are generalizable to a different population than the mentoring results.

b. Subclassification

Recall that the purpose of subclassification is to generate subclasses of units from each treatment group that are similar based on their covariates. This will allow for us to calculate the causal effect by simply taking the difference in potential outcomes and attributing the causal effect to the treatment assignment, assuming that we have all of the necessary covariates. Because we discarded the students for whom the unobservable summer surveys were used (as explained in Section a), we feel comfortable assuming that we have all of the covariates used to major advisor assignments for the students remaining in this component of the thesis.

I first separated students into their entering year cohorts and their STEM intended categories to examine the balance of covariates within those subclasses. Since they were not ideal, I then estimated propensity scores using a logistic regression on the treatment indicator with the following predictors: non STEM only indicator, STEM major 1 indicator, STEM major 2 indicator, ACT, SAT Math, SAT Verbal, SAT Writing, and Quantitative Reasoning Test scores, and entering years. These are some of the covariates in Tier 1, and so by putting them into the logistic regression function, I hope that balancing on the estimated propensity scores will establish balance on each of the covariates used to estimate them. Note that the underrepresented minority and first generation student indicators are also in Tier 1. To balance on these covariates, I explicitly calculate the proportion of students who are an underrepresented minority and those who are a first generation student in order to balance the proportions in both treatment groups. Furthermore, I considered the balance of students who were interested (by having an intended major) in Chemistry, Mathematics, and Neuroscience. I chose these three majors because they are amongst the most popular STEM majors at Wellesley.

After using propensity scores to create subclasses to refine the balance of covariates on intended majors and year, I noticed that the proportion of Biology intended majors were quite different in the two groups. Thus, I separated the data into students who had a Biology intended major and those who did not. After examining the balance of the covariates mentioned above, I decided to create subclasses to separate students who only had STEM intended majors and those who listed a STEM major first. I then further subclassified each subclass if there were many students in the subclass. If subclasses were too small, I combined subclasses when I could maintain the covariate balance and increase the sample sizes.

In total, there are fourteen subclasses. Table 10 contains a description of each subclass and the number of students who received active and control treatment within each subclass. Within each subclass, units of both treatment types can be compared to each other to determine a causal effect. However, before I can proceed, I must check that the covariates are balanced between the two treatment groups.

Table 10: Subclasses created to use for Advising component

	Subclass	Number of Students with STEM advisor	Number of Students with non STEM advisor
Student with interest in Biology and only declared STEM intended majors and has estimated propensity scores below the median.	1	19	14
Student with interest in Biology and only declared STEM intended majors and has estimated propensity scores above the median.	2	132	17
Students in entering class 2007 who declared all STEM intended majors but is not Biology.	3	45	11
Students in entering class 2008 who declared all STEM intended majors but is not Biology	4	24	13
Students in entering class 2012 who declared all STEM intended majors but is not Biology.	5	25	14
Students in entering classes 2008 - 2011 who declared all STEM intended majors but is not Biology.	6	99	15
Students who listed a STEM major first and has estimated propensity scores in the first quartile of propensity scores.	7	50	81
Students who listed a STEM major first and has estimated propensity scores in the first quartile of the second quartile of propensity scores	8	20	14
Students who listed a STEM major first and has estimated propensity scores in the second quartile of the second quartile of propensity scores	9	19	15
Students who listed a STEM major first and has estimated propensity scores in the third quartile of the second quartile of propensity scores	10	21	13
Students who listed a STEM major first and has estimated propensity scores in the fourth quartile of the second quartile of propensity scores	11	23	11
Students who listed a STEM major first and has estimated propensity scores in the third quartile of propensity scores	12	92	44
Students who listed a STEM major first and has estimated propensity scores in the first quartile of the fourth quartile of propensity scores	13	57	11
Students who listed a STEM major first and has estimated propensity scores in the second quartile of the fourth quartile of propensity scores	14	55	13

c. Covariate Balance

The goal of subclassification is to obtain a balance of covariate variables. Table 11 and Figure 2 show the balance of covariate variables after the subclassification. The left plot shows the actual differences in the variables whereas the right plot in Figure 2 shows the standardized differences. In both the table and the figure, the differences of the covariates between the active and control treatment groups are relatively small. In fact, in some cases, such as the indicator for STEM only intended majors, there is an exact match in the groups. Furthermore, Figure 3 shows a Love Plot of the standardized differences of the original data, after the non STEM only and STEM major 2 students have been omitted, and after subclassification. It is clear that subclassification led to better balanced covariates.

Table 11: Balance of Covariates After Subclassification

*Note that 980 students were included in the study but 13 students had missing treatment assignments

	Number of Students in Treatment	**Number of Students in Control**	**Weighted Mean Treatment**	**Weighted Mean Control**	**Diff in Wt Mean**
SAT Math	681	286	581.736	562.019	19.7173
SAT Verbal	681	286	569.337	555.191	14.1457
SAT Writing	681	286	575.823	563.805	12.0178
Biology indicator	681	286	.4002	.3373	.0629
Math indicator	681	286	.1641	.0668	.0973
Neuroscience Indicator	681	286	.1423	.2362	-.0939
Underrepresented minority	681	286	.1578	.2073	-.0496
First Generation	681	286	.0897	.1226	-.0329
ACT	681	286	11.3938	12.7722	-1.3784
QR score	681	286	13.5594	13.3575	.2019

Figure 2: Covariate Balance After Subclassification

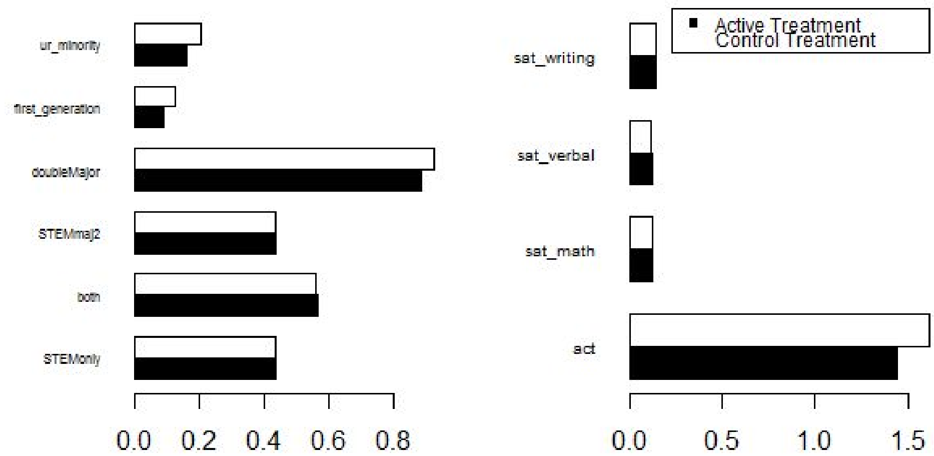
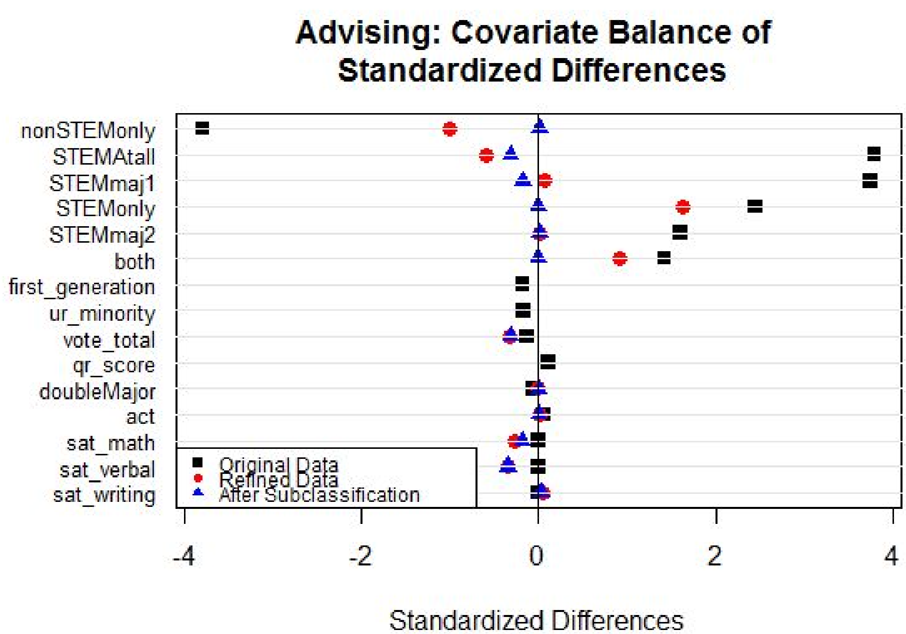


Figure 3: Love Plot Showing Covariate Balance n Original Data, Refined Data, and After Subclassification



d. Method for Analysis

Once covariate balance between both treatment groups is obtained, I estimated the effect of a STEM versus a non STEM advisor by calculating a weighted difference in means. The weights are calculated by $\frac{n_T + n_C}{N}$ where n_T is the number of students who received active treatment in the subclass and n_C is the number of students who received control treatment in the subclass. N is the total number of students in the advising component that are included in the analysis (which is 980). I then used these weights to calculate weighted mean differences and weighted variance of the differences. The weighted estimate and corresponding variance were used to generate 95% confidence intervals.

II. Results

The results are shown in Table 12.

Table 12: Weighted Difference in Potential Outcomes and Confidence Interval

	Weighted Mean Treatment	**Weighted Mean Control**	**Weighted Difference**	**Confidence Interval**
STEM at all indicator	.608	.581	.027	(-.049, .103)
STEM only indicator	.481	.438	.043	(-.035, .121)
Non STEM only indicator	.379	.406	-.027	(-.103, .049)
Cumulative GPA	3.32	3.35	-.029	(-.079, 022)

Looking at the weighted differences, all of them approximately zero. This suggests that the null hypothesis cannot be rejected. In addition, looking at the confidence intervals, all of them contain zero, which, again, suggests that the null hypothesis cannot be rejected for any of the potential outcomes. By not rejecting the null hypothesis, we are saying that having a STEM advisor does not mean that there is a higher chance that the student majors in STEM.

9. Discussion

I. Discussion of Overall Results

We did not find evidence that a student is more or less likely to major in STEM based on the STEM status of her first year mentor or first year advisor.

The null effect of the first year mentors is somewhat reasonable. Although students meet with their first year mentors almost every day during orientation, often times, this relationship ends after orientation week. Since most students do not meet with their first year mentors after the first week, it is possible that the mentors do not have an effect on the student's choice in major.

Similarly, in retrospect, we can seek to explain the null of effect of the first year advisors. Students are expected to meet with advisors at least once during the first week of classes. However, there are many times when this is the first and last time that a student meets with her first year advisor. One reason is that the faculty is not in the same major as the student's intended major. If this is so, she might not want to go to talk to her faculty advisor because they do not know anything about the major. Another reason is that the advisor and student are not proactive in setting up times to meet. Previous studies showed that the more times the student and faculty advisor meet, the better the benefits the students get (Campbell and Campbell, 1997). Note that data on how often students meet with their advisor is not recorded, and so we were not able to take that into account. One possible suggestion for administration, then, is to start collecting records of times students met with their advisors. Another possible explanation for the null effect in the advising component is that I only studied students who had known intended majors. Since these students have an idea of what they want to major in, it is possible that mentors and advisors do not have much influence because the students will not change their major choice. If we are able to obtain the summer surveys, then we would be able to study this idea.

Despite the lack of significance and small magnitudes in both the mentoring and advising component, the directions of observed differences in potential outcomes are as we might have hypothesized. The differences are positive for indicators of STEM only and STEM at all, which means that a student with a STEM mentor or advisor is slightly more likely to choose a STEM major. As for the non STEM only indicator and cumulative GPA, a student with a STEM mentor advisor is slightly less likely to major in a non STEM major and is more likely to have a lower GPA. This makes sense since STEM classes are often thought to be harder, and so if a student is more likely to major in a STEM major, she will have to take more STEM classes. This means that her GPA is more likely to decrease.

II. Discussion of Statistical Methods

There were many Statistical concepts in this thesis, some more conventional and some that are not as common. The Fisher Randomization Test and Separate Regressions Test were used to analyze the mentoring component of the thesis. These tests offered an alternative way to analyze the data, besides the typically used t-test, which would not be a good approach, given the complicated assignment mechanism used to assign mentors.

Within the mentoring component, the idea of a natural rerandomized experiment was also novel. Rerandomization is intended to be used prospectively, with the purpose of creating balanced groups in randomized experiments. However, in the mentoring component, there is a naturally occurring rerandomization of mentor group assignments if the person who was creating the mentor groups does not like the balance of the groups. This is interesting because although their goal of rerandomization is not for a hypothesis test, it is done in the same way as in a randomized experiment, and so I am able to analyze the mentoring data using randomization-based methods.

In the advising component, in order to analyze the data using Rubin's Causal Framework, subclassification was used to create subclasses that generated a balance of covariates. By using the Rubin Causal Framework to analyze the data, I am able to attribute the results to the fact that a student was assigned a STEM advisor. This is unlike many of the previous work that I looked at since the majority of them were observational studies, and you cannot draw causal inference from observational studies without using a causal framework.

III. Limitations

There are several limitations to this thesis that might have affected the results. One limitation was that the actual code used to create mentor groups was not available. I was only able to recreate the algorithm using the pseudo code given to me. However, because I was only given the general idea of the algorithm, I had to use my own judgment in dealing with some of the issues that occurred. One issue was that all the mentors' capacity of students were full even before all students were placed into a group. I just left these people out from the group because I was not explicitly told how the algorithm takes into account of these people, and I did not want to assume something that might be wrong.

Another limitation was that I was not given the summer forms that Dean os 2017 used to pair undeclared students and some remaining students with advisors. Because of this, undeclared students were discarded from the study. Omitting these undeclared students justifies our assumption of unconfoundedness. However, as mentioned in the previous section, this group is interesting to study because they are the students that are most easily influenced in terms of choice of major.

IV. Recommendations to Administration

Although we can come up with explanations in retrospect, our null results were a surprise to us and have implications for mentoring and advising. For the mentoring component, because mentoring groups are not formed to assist students with their major choices, a null effect is acceptable. Thus, the FYMs are doing their job, and there are no changes necessary to the current algorithm if the administration is not interested in forming groups to encourage a higher participation in a STEM major.

As for the advising component, there are two interpretations of the null effect. One interpretation is that the faculty advisors are not meeting with their students enough to influence what the student majors in. Another interpretation is that the faculty advisors are doing very well with their jobs and are able to provide students with information about majors that are not related to the advisor's department. If administration is interested in having the faculty advisors assist students in order to increase their enrollment in the advisor's respective field, then the current assignment process of the advisors and students does not seem effective. On the other hand, if administration does not care about increasing the enrollment in either the STEM majors or non STEM majors, the current assignment process does not need to be changed. This is because having a STEM advisor does not statistically significantly reduce the chance that student majors in a non STEM major. Thus, having a STEM (or non STEM) advisor does not "hurt" the student, and so the currently method of assignment by the Dean does not need to be changed. Perhaps, also, these results suggest that it is not worth putting too much time into carefully selecting advisor assignments, at least from the perspective of influencing STEM versus non STEM majors.

Lastly, I would recommend more data collection from both the mentors and advisors to have an idea of how often students are meeting with their mentors and advisors. With this information, it is possible to study the impact of more frequent contact on the student's choice of major.

10. Future Work

There are several other interesting factors from this thesis that can be further researched. One is attempt to access and to incorporate the summer forms that Dean of Class of 2017 uses to pair some of the advisors and students. Because I did not have these forms, undeclared students were excluded from the advising component. With the forms, I would be able to study more students. Secondly, an analysis on undeclared students might be interesting. Because these students are undeclared, they might be easier to influence, and so mentors and advisors might have a greater impact on the students. Another possible study is to define the active treatment of the advising component as the faculty advisor is in the same department as the student's intended major. If a student is placed with an advisor who is in the same department as her intended major, then the advisor can answer questions regarding the major. Furthermore, because of the common interest, they are more likely to meet after the initial meeting. Thus, students and faculty advisors will have greater contact, and the faculty advisors might have greater influence on the student. Another possible future work is to see whether there are small subsets of non STEM only or STEM major 2 students for whom we can actually draw causal inferences. Lastly, with more information collected about meetings between students and advisors, we can study the impact of contact and a student's choice of major.

On the statistical aspect, we could work to expand on David Watson's idea on creating Fisher intervals for binary values (Watson's slides on hand with me). We did not include this idea in this thesis as it is still a relatively new idea.

11. Conclusion

Overall, the magnitudes of the results were not practically significant, and they were rarely statistically significant. Thus, it is not possible to say that mentors and advisors influence a student's choice in major. Further analysis can be done in the future, specifically, on students who had undeclared intended majors who might be more easily influenced in terms of major choice. Furthermore, although the results were not as hypothesized, the statistical methods used to analyze the two components were interesting in the sense that this was one of the few research that I found that incorporated ideas of Causal Inference to determine the causal effect of mentors and advisors on choice of major. Additionally, we realized that this study includes a naturally rerandomized experiment which we analyzed accordingly- we believe that this approach is novel.

12. Appendix

Tier 1	Tier 2			Tier 3
act	sat_writing_missing	Theater Studies_indicator1	Music_indicator1	biracial
sat_math	sat_math_missing	Theater Studies_indicator2	Music_indicator2	black
sat_verbal	sat_verbal_missing	Theater Studies_indicatorAtAll	Music_indicatorAtAll	asian
sat_writing	act_missing	Neuroscience_indicator1	Computer Science_indicator1	latina
vote_total	qr_missing	Neuroscience_indicator2	Computer Science_indicator2	native_american
qr_score	satii_modern_hebrew	Neuroscience_indicatorAtAll	Computer Science_indicatorAtAll	white
STEMonly	satii_italian	Middle Eastern Studies_indicator1	Environmental Studies_indicator1	other_race
nonSTEMonly	satii_biology	Middle Eastern Studies_indicator2	Environmental Studies_indicator2	fy_cohort
both	satii_german	Middle Eastern Studies_indicatorAtAll	Environmental Studies_indicatorAtAll	foreign
STEMAtall	satii_german_with_listening	Chinese_indicator1	Anthropology_indicator1	
STEMmaj1	satii_japanese_with_listening	Chinese_indicator2	Anthropology_indicator2	
STEMmaj2	satii_korean_w_listening	Chinese_indicatorAtAll	Anthropology_indicatorAtAll	
doubleMajor	satii_french_with_listening	Classical Civilizations_indicator1	Cognitive&Linguistics Sc_indicator1	
Ur_minority	satii_latin	Classical Civilizations_indicator2	Cognitive&Linguistics Sc_indicator2	
first_generation	satii_chinese_with_listening	Classical Civilizations_indicatorAtAll	Cognitive&Linguistics Sc_indicatorAtAll	
	satii_spanish_with_listening	Architecture_indicator1	Media Arts and Science_indicator1	
	satii_world_history	Architecture_indicator2	Media Arts and Science_indicator2	
	satii_french	Architecture_indicatorAtAll	Media Arts and Science_indicatorAtAll	
	satii_spanish	Chemistry_indicator1	Russian_indicator1	
	satii_eco_biology	Chemistry_indicator2	Russian_indicator2	
	satii_physics	Chemistry_indicatorAtAll	Russian_indicatorAtAll	
	satii_molecular_biology	Clscl&Nr Eastern Archeol_indicator1	Studio Art_indicator1	
	satii_mathematics_level_i	Clscl&Nr Eastern Archeol_indicator2	Studio Art_indicator2	
	satii_chemistry	Clscl&Nr Eastern Archeol_indicatorAtAll	Studio Art_indicatorAtAll	
	satii__us_history	Astrophysics_indicator1	Russian Area Studies_indicator1	
	satii_literature	Astrophysics_indicator2	Russian Area Studies_indicator2	
	satii_mathematics_level_ii	Astrophysics_indicatorAtAll	Russian Area Studies_indicatorAtAll	
	Biological Sciences_indicator1	Japanese_indicator1		
	Biological Sciences_indicator2	Japanese_indicator2		
	Biological Sciences_indicatorAtAll	Japanese_indicatorAtAll		
	Internat'l Relations_indicator1	Compar Literature_indicator1		
	Internat'l Relations_indicator2	Compar Literature_indicator2		
	Internat'l Relations_indicatorAtAll			
	History_indicator1			
	History_indicator2			
	History_indicatorAtAll			
	Economics_indicator1			
	Economics_indicator2			
	Economics_indicatorAtAll			

	Undeclared_indicator1 Undeclared_indicator2 Undeclared_indicatorAtAll English_indicator1 English_indicator2 English_indicatorAtAll Political Science_indicator1 Political Science_indicator2 Political Science_indicatorAtAll Psychology_indicator1 Psychology_indicator2 Psychology_indicatorAtAll French_indicator1 French_indicator2 French_indicatorAtAll Art History_indicator1 Art History_indicator2 Art History_indicatorAtAll Philosophy_indicator1 Philosophy_indicator2 Philosophy_indicatorAtAll East Asian Studies_indicator1 East Asian Studies_indicator2 East Asian Studies_indicatorAtAll Africana Studies_indicator1 Africana Studies_indicator2 Africana Studies_indicatorAtAll Compar Literature_indicatorAtAll Peace & Justice Studies_indicator1 Peace & Justice Studies_indicator2 Peace & Justice Studies_indicatorAtAll Biological Chemistry_indicator1 Biological Chemistry_indicator2 Biological Chemistry_indicatorAtAll Women's Studies_indicator1 Women's Studies_indicator2 Women's Studies_indicatorAtAll French Cultural Studies_indicator1 French Cultural Studies_indicator2 French Cultural Studies_indicatorAtAll Sociology_indicator1 Sociology_indicator2 Sociology_indicatorAtAll Latin_indicator1 Latin_indicator2 Latin_indicatorAtAll Spanish_indicator1 Spanish_indicator2 Spanish_indicatorAtAll Italian Studies_indicator1 Italian Studies_indicator2 Italian Studies_indicatorAtAll German Language & Lit._indicator1 German Language & Lit._indicator2 German Language & Lit._indicatorAtAll Religion_indicator1 Religion_indicator2 Religion_indicatorAtAll American Studies_indicator1 American Studies_indicator2 American Studies_indicatorAtAll Jewish Studies_indicator1 Jewish Studies_indicator2 Jewish Studies_indicatorAtAll Greek_indicator1 Greek_indicator2 Greek_indicatorAtAll Medieval & Renaissance_indicator1 Medieval & Renaissance_indicator2 Medieval & Renaissance_indicatorAtAll Geological Sciences_indicator1 Geological Sciences_indicator2 Geological Sciences_indicatorAtAll Latin American Studies_indicator1 Latin American Studies_indicator2 Physics_indicator1 Physics_indicator2 Physics_indicatorAtAll Mathematics_indicator1 Mathematics_indicator2 Mathematics_indicatorAtAll Cinema & Media Studies_indicator1 Cinema & Media Studies_indicator2 Cinema & Media Studies_indicatorAtAll	

13. Works Cited

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity‐score matched samples. *Statistics in medicine*, *28*(25), 3083-3107.

Beggs, J. M., Bantham, J. H., & Taylor, S. (2008). DISTINGUISHING THE FACTORS INFLUENCING

Belson, W. A. (1956). A technique for studying the effects of a television broadcast. *Applied Statistics*, 195-202. COLLEGE STUDENTS'CHOICE OF MAJOR. *College Student Journal*, *42*(2), 381.

Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational behavior*, *23*(3), 329-345.

Campbell, T. A., & Campbell, D. E. (1997). Faculty/student mentor program: Effects on academic performance and retention. *Research in higher education*, *38*(6), 727-742.

CCAP Grading Policy FAQ. (2012, April 19). Retrieved April 17, 2017, from http://www.wellesley.edu/registrar/grading/gradingpolicyfaq#3xAAfG1SWUuuk54m.97

Crisp, G., Nora, A., & Taggart, A. (2009). Student Characteristics, Pre-College, College, and Environmental Factors as Predictors of Majoring in and Earning a STEM Degree: An Analysis of Students Attending a Hispanic Serving Institution. *American Educational Research Journal,46*(4), 924-942. doi:10.3102/0002831209349460

First-Year Advising. (2016). Retrieved April 17, 2017, from http://www.wellesley.edu/esp/entering/academics/advising

First-Year/Transfer Mentor. (2016). Retrieved April 16, 2017, from http://www.wellesley.edu/reslife/student_leader/positions/first-year-transfer-mentor#g7qRgz8S4JlC4VhR.97

Fisher, R. A. (1925). Statistical Methods for Research Workers. *Oliver and Boyd*. (Ed.1).

Fisher, R. A. (1935). Design of Experiments. *Oliver and Boyd.*

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945-960.

Imbens, G., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: an introduction*. New York: Cambridge University Press.

Morgan, K. L., & Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 1263-1282.

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

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in medicine*, *26*(1), 20-36.

Sciences @ Wellesley. (2016). Retrieved April 17, 2017 from *Wellesley College*. N.p., n.d. Web. 06 Dec. 2016.

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), 464-474.

Wellesley Facts. (2016). Retrieved April 16, 2017, from http://www.wellesley.edu/about/wellesleyfacts#yWBuJq5eg8esoulB.97