A TRIDENT SCHOLAR
PROJECT REPORT

NO. 481

Soft Skills and Soft Standards in a Sequential Learning Framework

by

Midshipman 1/C Megan L. Hanson, USN

UNITED STATES NAVAL ACADEMY
ANNAPOLIS, MARYLAND

USNA-1531-2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
5-20-19

2. REPORT TYPE

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Soft Skills and Soft Standards in a Sequential Learning Framework

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hanson, Megan L.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Naval Academy
Annapolis, MD 21402

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Trident Scholar Report no. 481 (2019)

12. DISTRIBUTION / AVAILABILITY STATEMENT

This document has been approved for public release; its distribution is UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This paper studies the effect of soft skills in higher education by analyzing how personality types and instructor tendencies affect student performance through six sets of sequential classes. To do this, we look at freshman students at the United States Naval Academy from the class years of 1998 to 2018. The Naval Academy offers an ideal environment to test the effects of soft skills due to the unique controls of the academic environment: all students take the same core courses their freshman year and are randomly placed in sections with no ability to select instructors or peers. We analyze instructor grading tendencies in a grade-distortion model while controlling for a variety of background characteristics and accounting for student personality types as captured by Myers-Briggs. We define "cushy" instructors as those who, on average, tend to give higher grades than students are projected to receive based on background characteristics. Similarly, we define "challenging" instructors as those who tend to give lower grades than we would expect a particular student to achieve.
We find that teachers have the most significant effect on subsequent student performance. Excessively "cushy" instructors hurt student performance in follow-on courses, especially in STEM courses. We also find that student personality measures matter for academic achievement overall and in a course-by-course analysis: most notably, we see that "judgers" outperform "perceivers" across all courses and lead to a higher overall grade point average over all four years. Finally, we see that the gender of the student and instructor may play a role in the ability of a student to succeed in a follow-on course regardless of the signal an instructor sends with a grade.

15. SUBJECT TERMS
education, instructor effects, personality data, human capital

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		63	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

U.S.N.A. --- Trident Scholar project report; no. 481 (2019)

SOFT SKILLS AND SOFT STANDARDS IN A SEQUENTIAL LEARNING FRAMEWORK

by

Midshipman 1/C Megan L. Hanson
United States Naval Academy
Annapolis, Maryland

____
(signature/date)

Certification of Adviser(s) Approval

Associate Professor Ahmed S. Rahman
Economics Department

____
(signature/date)

Professor Katherine A. Smith
Economics Department

____
(signature/date)

Assistant Professor Alexander F. McQuoid
Economics Department

____
(signature/date)

Acceptance for the Trident Scholar Committee

Professor Maria J. Schroeder
Associate Director of Midshipman Research

____
(signature/date)

USNA-1531-2

Abstract:

This paper studies the effect of soft skills in higher education by analyzing how personality types and instructor tendencies affect student performance through six sets of sequential classes. To do this, we look at freshman students at the United States Naval Academy from the class years of 1998 to 2018. The Naval Academy offers an ideal environment to test the effects of soft skills due to the unique controls of the academic environment: all students take the same core courses their freshman year and are randomly placed in sections with no ability to select instructors or peers. We analyze instructor grading tendencies in a grade-distortion model while controlling for a variety of background characteristics and accounting for student personality types as captured by Myers-Briggs. We define "cushy" instructors as those who, on average, tend to give higher grades than students are projected to receive based on background characteristics. Similarly, we define "challenging" instructors as those who tend to give lower grades than we would expect a particular student to achieve.

We find that teachers have the most significant effect on subsequent student performance. Excessively "cushy" instructors hurt student performance in follow-on courses, especially in STEM courses. We also find that student personality measures matter for academic achievement overall and in a course-by-course analysis: most notably, we see that "judgers" outperform "perceivers" across all courses and lead to a higher overall grade point average over all four years. Finally, we see that the gender of the student and instructor may play a role in the ability of a student to succeed in a follow-on course regardless of the signal an instructor sends with a grade.

Keywords: education, instructor effects, personality data, human capital

Acknowledgements:

This project has been a yearlong endeavor and could not have been done without the help of the many advisors, professors, and students over the last year. Thank you to the Trident Committee and Professor Schroeder for the opportunity to pursue this research and for your guidance and support. Thank you to my advisers for their dedication, hard work, patience, and standards (of which we learned a lot about). Thank you to the Economics Department for always providing guidance and for answering my Stata questions if I appeared in your office lost and distraught. Thank you to all of the many people around the yard I had read parts of my paper to see if they could understand what I truly meant by “soft skills”. Thank you to my teammates, roommates, and friends for listening to the constant pounding of keys and providing all of the emotional support required to do this project. I could not have done it without you.

HRPP Approval:

U.S. Naval Academy Human Research Protection Program
Nimitz Library G10 - Mail Stop 10M - Annapolis, Maryland 21402

MEMORANDUM

20 Apr 18

From: Ms. Erin Johnson, USNA HRPP Office
To: MIDN Megan Hanson, Mathematics Department

Subject: APPROVAL OF HUMAN SUBJECT RESEARCH

Ref: (a) SECNAVINST 3900.39E
(b) 32 CFR 219
(c) USNA HRPP Policy Manual

USNA Assurance # DoD N-40052

HRPP Approval # **USNA.2018.0019-IR-EP7-A**

1. The Superintendent, as the Institutional Official (IO), approved your research protocol "Teachers, Timing, and Tenacity" involving human subjects.
   a. The USNA approval date is 13 April 2018. The approval is in effect for one year.
   b. The IRB Chair reviewed and recommended approval on 10 April 2018.
   c. The protocol approval is in effect for one year from the recommended approval date of the IRB Chair and will expire on 9 April 2019.
   d. If the research is to be continued beyond 9 April 2019, please submit your renewal application to this office by 9 March 2019 to allow time for adequate processing.

2. The research is determined to be of minimal risk and was reviewed under expedited review according to reference (b) as research on individual or group characteristics employing survey/interview methodologies (Category 7).

3. Per the USNA HRPP Policy and Procedures manual (Section IX), if there should be any changes to your research, you must submit an amendment immediately to the HRPP office to process the revisions and secure approval of the IO.

4. You are required to report when the research has concluded according to Section XIII of the USNA HRPP Policy and Procedures manual and to provide this office with copies of any articles or presentations resulting from this research. Additionally, any presentations or publications must include acknowledgment of IRB approval using the HRPP approval number.

5. If you have any questions, please contact this office at 410-293-2533 or HRPPoffice@usna.edu.

ERIN JOHNSON
USNA HRPP Office

DEPARTMENT OF THE NAVY
UNITED STATES NAVAL ACADEMY
121 BLAKE ROAD
ANNAPOLIS MARYLAND 21402-1300

3900
10 Apr 18

MEMORANDUM

From: Chair, Institutional Review Board (Code 28)
To: Superintendent, U.S. Naval Academy

Subj: HUMAN SUBJECT RESEARCH BY MIDN MEGAN HANSON (MATHEMATICS DEPARTMENT)

Ref: (a) SECNAVINST 3900.39E
(b) 32 CFR 219
(c) USNA HRPP Policy Manual

Encl: (1) Protocol Package for MIDN Megan Hanson (Forms 2, 3, 4, 5, 5A, CITI, and Supplemental Information)

1. Per references (a) through (c), I have reviewed the research protocol for initial review submitted by MIDN Megan Hanson from the Mathematics Department titled "Teachers, Timing, and Tenacity." The co-investigator is Professor Ahmed Rahman from the Economics Department.

2. The purpose of this research is to look into how Midshipmen optimize their effort level based on their natural aptitude, their professor and their personality. All data has been de-identified.

3. The research is determined to be of minimal risk and approval is recommended for one year. The research is considered expedited according to reference (b) as research on individual or group characteristics (Category 7).

10 APRIL 18
MICHAEL R. KELLERMANN

Date: 4/13/18

☑ Approved ☐ Modification ☐ Disapproved

Comments:

W. E. CARTER, JR
Vice Admiral, U.S. Navy
Superintendent

Table of Contents:

I. Introduction…………………………………………………………………………6
II. Previous Literature……………………………………………………………….....10
III. Methodology………………………………………………………….............16
   a. Soft Skills and Educational Outcomes…………………………………….....…17
   b. Soft Standards…………………..………………………………………….….18
   c. Soft Skills and Soft Standards Effects on Educational Outcomes ……….……...21
IV. Data……………………………………………………...………...................24
   a. Student Data…………………….…………………...………...................24
   b. Personality Data……………………………………...………...................28
   c. Faculty Data……………………………………...……….......................31
V. Results…………………………………………………………………………..33
   a. The Effect of Soft Skills on Overall Student Performance …..…………………..33
   b. Soft vs Hard Standards Index by Faculty in a Sequential Learning Framework...37
   c. Soft Skills and Soft Standards…………………………………………………...41
   d. Gender and Outcomes…………………………………………………………48
VI. Conclusion……………………………………………………………………...53
VII. References…………………………………………………………………………55
VIII. Appendix…………………………………………………………………...…...57

I: Introduction

The majority of academic research on education has focused on academic success as measured by "hard skills". In secondary education, student scores on statewide-standardized tests often measure teacher effectiveness while grade point averages and test scores measure student achievement. These hard skills are how students are ranked in high school and these are the metrics that institutions of higher learning use to admit their students, but the potential for success is being mismeasured if soft skills matter for one's ability to develop human capital and instructor standards play a role in student success. We define soft skills as personality traits and intangible qualities that complement hard skills, which encapsulate test scores, GPA, and other background information. We define soft standards as instructor tendencies as captured by their grading tendencies: whether they tend to award grades higher or lower than we would expect a given student to achieve. This project is an investigation into the degree to which student soft skills and instructor standards contribute to the efficacy of education at the college level.

As universities seek to develop their students to their full potential, soft skills may be an overlooked important input to education. Much of the current research into higher education implies a production function of student success, but they omit any soft skills of the students or instructors involved. Becker (1964) proposes that human capital is a function, with effort and innate abilities as important inputs to the production of human capital[1]. We see two missing factors from Becker's production function: student soft skills (their ability to rise to a challenge)

[1] "Gary Becker's Concept of Human Capital." The Economist, The Economist Newspaper, 3 Aug. 2017.

and faculty member expectations (pushing students to achieve), both of which we analyze and their respective impacts, both joint and individually on student success.

For students, we proxy for soft skills by using their personality types as measured by the Myers-Briggs personality test. As we analyze the effect of personality types on academic achievement, we look at students' specific personality traits and characteristics within the larger personality type. For example, the most common personality type of students in this population is ESTJ (Extraversion, Sensing, Thinking, Judging), but we individually pull out personality effects. For example, we analyze the effect of a student being a "judger" and the impact that has across all courses.

For instructors, we proxy for their soft standards by looking at their grading tendencies across all students (minus the student being examined). By comparing what a student is projected to achieve in a course versus the grade they are awarded by the instructor and summing these differences across all students of a particular professor, we are able to create a picture of instructor tendencies. From this, we can classify an instructor as "cushy", with a tendency to give higher grades than we initially predict the student to achieve, or "harsh", with a tendency to give lower grades than we predicted the student would achieve based on background characteristics.

Analysis of the importance of soft skills within higher education has proven to be a problem for decades of educational research in part due to lack of data. Most research that does account for soft skills of students or professors is often limited in size or scope because of the necessary data to do such a study[2]. For these reasons, the United States Naval Academy (USNA)

[2] This is discussed further in previous research, but personality studies tend to be limited in size and scope because students must all take the same personality test in order to use this as a factor, which requires time, money, and the availability of data of student scores.

proves to be an ideal place to conduct this research. USNA has extensive data for each of its students: it includes a student's final grades received, the instructor for each class, a student's background statistics, and student personality types for the entire population (approximately 40,000 observations). Because we have extensive classroom and grades data, we are afforded the opportunity to analyze trends across a range of classes that we feel present a good academic cross-section of introductory courses in both STEM and humanities.

We look at sequential learning and see the effects of introductory teachers on student performance in follow-on classes across six sets of classes. The Naval Academy has an extensive core curriculum that all student take, with subjects following a sequential curriculum across two semesters. For example, a freshman will take Calculus I, English I, Chemistry I their first semester and go on to take Calculus II, English II, and Chemistry II their second semester. The Naval Academy records extensive data on student performance as well as background academic and demographic information. Access to this data allows us to analyze the effects of personality characteristics of students and teachers across years, genders, and a range of courses.

This study is unique because our two-step methodology presents the effects of personality traits on student courses across six classes, but it also allows us to see trends across STEM and humanities courses, across genders of teachers and students in order to see which groups are most heavily impacted, and larger trends that we see in student performance overall. The conclusions drawn from this study can be applied to other universities, but in a broader sense, this study helps quantify the importance of soft skills in academic success for students at all education levels and instructor quality in student learning.

This paper is organized into the following sections: a synopsis of relevant literature, an explanation of methodology, a presentation of the data, a discussion of results, and a conclusion that offers suggestions for further research.

II. Previous Empirical Research

This research is addressing the extent to which the soft skills of students and the soft standards of instructors impact academic success. Previous research of the effect of soft skills on student achievement comes in several tenets[3]. First, some empirical research looks at whether personality type matters for success in higher education in small samples[4]. More specifically, there is research done at the university level to determine if specific personality traits matter for academic achievement and how student personality types interact with those of their professors. However, this research is usually limited to a very small sample of students and it only considers a single class, leaving room for many omitted variables. There is another vein of research that focuses on faculty soft skills, including personality traits and "methods" of teaching. This research is unique because it couples all of these areas of research: we investigate the soft skills of both students and faculty and their interaction and impact upon student achievement in a large sample size across two decades of students and across six sets of sequential courses.

Research indicates personality types may affect academic success in higher education while controlling for other factors in several limited studies. Lundberg et al. (2013) research the effects of personality types on college graduation; they find that level of academic achievement is not tied to a specific personality trait; rather, the traits that lead to success differ based on student socioeconomic status and environment. Despite research on personality and education,

---

[3] Other papers dot necessarily use the term "soft skills", but they research personality and teacher efficacy (through standards), which we define in this paper as soft skills which is why it is included and referred to as such
[4] Sample size is normally less than 100 students.

there is a gap in the literature that considers the personality implications for a large sample of students within higher education.

Heckman et al (2012) correlate success more highly with certain personality traits than with IQ. Aspects of the personality, specifically *Conscientiousness* from the OCEAN Model (one of several highly used personality tests), served as a more accurate predictor for success than any other facet of the intelligence test. While skills or intelligence are often directed towards a certain area, *Conscientiousness* applied more broadly to success in jobs and in academia because *Conscientiousness* does not vary as a predictor for success between jobs or job complexity. Heckman et al (2012) estimated their findings based on a sequential model of learning in order to best study personality traits, which is the same overarching model that we use in this study. While this research includes *Conscientiousness* from the OCEAN model rather than another personality test, it echoes research done using the Myers-Briggs test (MBTI) which is a similar personality test[5]. This research provides value-add in a larger sample size across a longer time-horizon, a different personality test, and the inclusion of more than a single dimension of the personality test.

David Keirsey and Marilyn Bates developed four temperaments of students (focusing on secondary and post-secondary education) based on combination of MBTI categories (SP, SJ, NT, NF)[6]. Most important to note in this research is the SJ combination (called Epimethean). Epimetheans tend to love “rules, regulation, duty, and honor and have a strong work ethic and parental outlook”. Much of the sample in this research is SJ (as the dominant personality type at

---

[5] Both tests use a scaling system for certain personality traits. The OCEAN Model uses 5 personality traits on a continuous scale whereas the Myers-Briggs test uses four sets of personality pairs for comparison of difference traits.

[6] Keirsey, D., and M. Bates. 1984. Please understand me: Character and temperament types. 5th ed. Del Mar, Calif.: Prometheus Nemesis.

the Naval Academy is ESTJ), so this distinction is important to note. Kiersey and Bates found that the SJ temperament is normally overrepresented in university students and in K-12 educators (although this personality type only makes up 38% of the general population), but SJ personality types are underrepresented by most university professors (Myers and Mcaulley, 1985). Most university professors tend to be NF or NT, which, they found, could be why certain students who excelled in high school suddenly struggle to connect with their teachers or struggle with a change in learning style. A table of the four temperaments for students and teachers is provided in the appendix.

Emerson (2007) found that gender and personality are often linked. Emerson (2007) ran a regression to understand the relationship between personality type as captured by Myers-Briggs and academic achievement while accounting for demographic factors (gender being one of them). Typically, researchers found that females tended to underperform compared to their male counterparts, but when variables were added for personality traits, the gender differences decreased, meaning some of the effects of gender may be attributed to personality rather than purely gender but there may be personality types more correlated with females or males. This research builds upon the research of Emerson with a larger sample size and a more diverse range of classes and the ability to look at the effects of gender in a sequential learning framework.

A. Instructor Effects on Higher Education

The effect of students and teachers temperaments has also been found to affect academic achievement in small samples. Using the population of a Principles of Microeconomics class, Borg and Shapiro (1996) found that personality types of students and teachers and the degree to which they match affect student performance. Students with the SJ personality type (dependent learners) and introverts tended to perform better in this class because it is taught in a similar way to high school courses (with lecture-based material). This is worth noting because their results differ from another paper that found that independent learners (NF or NT temperaments) tended to do best in introductory economics courses, although this could be due to the closer temperament match with their instructor[7]. Research indicates that the matching of student and instructor temperaments matters for academic achievement and that when the two temperaments match, the students tends to perform better. Our research builds upon this by considering a much larger population and considering six sets of classes, allowing us to broaden our aperture and consider the effects of the interactions of soft skills across subject material and sequential semesters.

Beyond personality and its correlation to academic achievement in certain classes, other faculty effects in higher education has been explored[8]. Much of the results were mixed on the effects of observable characteristics and background information on instructors, such as evaluations (Jacobs and Lefgren, 2004), National Board Certification and teacher licensure (Clotfelter, Ladd, Vigdor, 2006, 2007), gender and race of student and professor (Dee, 2004, 2005). The general findings are that characteristics alone cannot determine teacher quality as no observable characteristic has been found to be deterministic and repeatable across populations; research finds vast differences of instructor effects across schools and even within schools[9]. The outcomes that result from teacher quality are substantial: there are instances where a particular teacher can cover 1.5 years of material in a single year whereas other teachers can only cover half a year. If a student is caught by two bad years of teachers, they may be unable to recover,

[7] Wetzel, Potter, O'Toole (1982)
[8] Including both hard and soft skills, discussed further in literature section.
[9] A summary of value-add studies of instructors on student achievement can be found in Hanushek and Rivkin (2010).

especially when compared to a student who has had two good years of teachers[10]. A consensus emerges, however, that there is no other attribute in higher education that is more influential than teachers are, which is why research continues to be done trying to determine what exactly makes a teacher most helpful for student achievement[11].

In one such paper investigating observable professor characteristics done at the Air Force Academy, Carrell et al. (2010) investigate the effects of teacher quality on student learning, specifically using student performance in two classes and evaluations[12]. This paper looked at the series of classes, Calculus I and Calculus II, in order to determine how students reward or punish their teachers for deep learning in their student evaluations. This project illuminates how teacher quality makes a difference in the depth of the learning incurred in a class, measured by how well a student performs in the follow-on class, and how students then rate their teachers based on this outcome. Our research improves upon this by considering more than a single class (we use six sets of sequential courses). Our research differs in how we account for faculty characteristics. While we also use the standards of instructors as one of our considerations, we look at observable differences in standards through grading practices rather than student evaluations.

Looking beyond observable characteristics of professors, Wetzel, Potter, O'Toole (1982) and Charking, O'Toole, Wetzel (1982) investigated effect of learning styles of the professor and how it relates to the learning style of the student. The research found that, in general, the more closely a student's temperament matches their instructors, the better they perform in the class.

---

[10] Hanushek, Eric A. "Economic Aspects of the Demand for Teacher Quality." *Economics of Education Review*(2010).
[11] Typical assumptions of environments or achievements that may make a teacher more productive are master's degree, classroom size, trainings, etc.
[12] Also called student opinion forms

These results complement this study as we also look into the effects of the soft skills of teachers, but rather than looking through the lens of teacher personality, we choose to look at teacher grading tendencies and the effect that these tendencies have on the performance of students.

In a paper investigating the effect of instructor gender on student performance, Carrell et al. (2010) find that the gender of introductory course instructors have an effect on student performance in STEM courses. The group most impacted by the gender of their instructors is high-achieving female students in these STEM courses while male students have no difference in effect when they have a male versus a female instructor. Carrell et al. find that the gender gap is nearly eliminated when female students have female instructors and they find that more females tend to pursue STEM majors when they have a female instructor for their introductory course. The results from this paper led us to consider the effect of instructor gender, but we are able to pursue a more long-term approach and see the effect on academic achievement beyond the initial course.

The value added in this research comes from the combination of three major veins of previous literature: soft skills and their impact on academic achievement, the effect of gender on higher education and academic achievement, and the effect of instructor "quality". This research also provides a larger sample size of students and instructors than has been available in previous research on the subject, it spans two decades, and it follows six sets of sequential classes across a range of subjects included both STEM and non-STEM courses. This research provides a comprehensive and qualitative look at the effects of soft skills and soft standards on learning.

III: Methodology

This paper is an investigation into soft skills, soft standards, and how these two interact within the classroom to develop human capital in higher education. First, we consider the soft skills of students as captured by the Myers-Briggs personality test and investigate how student personality traits or combinations affect student performance. We look beyond the soft skills of students though and extend the analysis to instructors. We consider faculty soft skills through a different lens: rather than their personality types, we consider their observable tendencies in the classroom. More specifically, we look at how these soft skills manifest in the standards that instructors enforce in their grades, and we categorize instructors as either "cushy" or "harsh" depending on these tendencies.

With the knowledge that both the instructor's and the student's soft skills affect educational success, we look to the interaction of soft skills and soft standards in a classroom environment across a range of subjects. In research that centers on soft skills, there has historically been some degree of selection bias and there is rarely the possibility of a random sample. The Naval Academy presents a unique environment for investigating soft skills and soft standards for several reasons, notably the random assignment of peers and professors, a common core set of courses, and availability of background characteristics of students.

When students are admitted to the Naval Academy, they are randomly placed in one of thirty companies, which becomes their immediate peer group as students will room, eat, and attend all events with their companies. Freshman at the Naval Academy offer a particularly useful population to study because of the similarity in their academic course load. Students all take the same core curriculum their first year and they are randomly assigned to sections and professors, there is no opportunity for any student input on preferences for professor, time of day, or section.

All freshman at the Naval Academy have a common set of core courses across a range of subjects. Specifically, there are five sets of sequential courses during a student's freshman year over which we study student achievement (Calculus, English, Chemistry, Government, and History). We also study one set of courses over a student's sophomore year: Physics. This common set of core courses allows us to study student achievement across a range of courses irrespective of major selection, attrition bias, and selection bias. More specifics on the courses studied are discussed in the data section.

The Naval Academy presents a unique experience for academic research because of the data that is available for each student. The Naval Academy ensures that its student body is a geographically and socioeconomically diverse group that increases in diversity measures each year. The data that allows this research in particular to be conducted is the information that the Naval Academy has on the personality types of its students. Every student takes the Myers-Briggs (MBTI) personality test their first year so we have data from an entire population that presents an opportunity to account for some of these soft skills as measured by personality.

We measure student outcome by looking at overall GPA in a student's four years as well as more specifically in a sequential course-by-course basis and take into account student personality, teacher characteristics, and the interaction of student personalities with teacher tendencies.

A. Soft Skills and Educational Outcomes

We start by considering the impact of student personality on college success, which has been largely absent in the literature surrounding higher education and student success. We isolate the impacts from soft skills by including them as factors influencing student success, accounting for them the same way we do for harder skills, such as academic background and demographic characteristics. In order to proxy for these soft skills, we use the Myers-Briggs personality test that each student takes during their freshman year. We consider the following empirical model.

Equation 1: OLS Equation for Student Performance

$$GPA_i = \alpha + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \gamma_3 X_{3i} + \varepsilon_i \qquad (1)$$

where:

GPA_{ij}: Overall GPA earned by student *i* over college career

X_{1i} : Matrix of demographic characteristics of student *i*

- Gender of student
- Minority status
- Recruited athlete status

X_{2i} : background academic characteristics of student *i*.

- SAT Verbal and Math scores
- weighted high school rank

X_{3i} : personality characteristics of student *i*[13].

- Extraversion-Introversion Intensity Index
- Sensing-Intuition Intensity Index
- Thinking-Feeling Intensity Index
- Judging-Perceiving Intensity Index

B. Soft Standards

As we further unpack the production function for human capital accumulation, we next consider the impact of faculty standards. We start by estimating a faculty-specific impact, which is included as an additional explanatory factor in student learning over time. In the first stage, we "back out" the instructor effects after controlling for everything that we know affects student

[13] Personality characteristics are measured by an intensity index of the four dichotomous traits of the Myers-Briggs personality test. More specifics of the test will be given in the data section.

performance, including hard skills (SAT scores in math and verbal, a weighted high school rank, gender, minority status, whether they are a recruited athlete, feeder source if not a direct admit (NAPS, Boost, Foundation, Prior Enlisted Nuclear), as well as soft skills (personality intensity indices). In order to do this, we regress the fall grade on all of the background characteristics and focus on the residuals measuring the difference between a student's expected grade in a course and the actual grade received in a course[14]. The summation of student residuals are then calculated for a specific faculty member to form an instructor-specific measure of "grade distortion". Because of the random assignment of students and professors, the probability of having over performing and underperforming students will average out, thus leaving an instructor grade-distortion effect. This regression is run for each student, *i*, in each course, *j*.

Equation 2: First-Stage Estimation

$$GF_{ij} = \alpha + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \gamma_3 X_{3i} + \varepsilon_{ij} \qquad (2)$$

where:

GF_{ij}: Grade earned by student *i* in Fall course.
X_{1i} : demographic characteristics of student *i* (estimates not displayed here).
X_{2i} : background academic characteristics of student *i*.
X_{3i} : personality characteristics of student *i*.

[14] The fall grade is the final grade that a student receives in their first class in a sequential model. For our purposes, we refer to Class 1 in the model as the fall class and Class 2 in the sequential learning model as the spring class. For example, Calculus I would be the fall class and Calculus II would be the spring class and only the final grade from each is used for the model.

This regression yields several important pieces of information. First, the values for γ_3 give information about how much specific personality traits affect the grade. For example, if γ_3 for the Sensing-Intuition intensity index was 0.005, if a student were to change from being a 0 in Sensing (essentially neutral between intuition and sensing) to a 100 in Sensing, their GPA would increase by 0.5 in the class (half of a letter grade).

From this regression, we obtain the average distortion for the entire course (for example, the average grade distortion across all students that have ever been taught Calculus I). We then take this residual and sort it by each unique instructor and aggregate the average grade distortion for a particular instructor, this capturing their individual effect.

Equation 3: Residual Aggregation

$$\Omega_k^{fall} = \sum_{ij=1}^{k} \varepsilon_{ij} \qquad (3)$$

for the N students that instructor k has taught in the relevant course. The summation is done for all students that a professor has taught, excluding the student that is under consideration in the particular regression.

This average residual shows a grading trend for each instructor over all students. For example, if an instructor has a high positive average residual, she or he may be inflating grades, teaching to the test, be less likely to give students low grades, or he is a good instructor who is teaching the students well enough that they all earn higher grades than they were originally predicted to achieve. Conversely, if an instructor has a negative residual, she or he may be deflating grades, focusing on deeper learning rather than teaching to the test, be less likely to award high grades to a student, or the students may just not be learning as effectively in the class. While the precise mechanism is unobserved, the net result is that certain faculty provide unexpectedly high or low grades, and professors are effectively randomly assigned to students with no choice over course selection, meaning we incur little to no selection bias.

C. *Soft Skills and Soft Standards Effects on Educational Attainment*

Next, we turn to the spring course outcome for these students. All of the courses we analyze follow a sequential learning framework with a follow-on course in the second semester. The timing of courses allows us to track the continuity of learning in the same subject and across the range of classes that freshmen take at the Naval Academy. For example, if a student takes Calculus I, English I, and Chemistry I in the fall they will almost invariably take Calculus II, English II, and Chemistry II in the spring. The approach of studying multiple classes allows us to see that sequential learning may be different in the Calculus I/Calculus II framework than it is in the History/Government framework, especially as we split down the STEM vs non-STEM course lines.

The second stage regresses a student's spring grade on their fall grade and background characteristics as before, but this regression also includes the mean residual of a student's fall instructor as well as the mean residual of the student's spring instructor. This regression is run for each student, i, in each course, j, with a specific instructor, k.

The regression is shown below:

Equation 4: Impacts of Soft Standards on Sequential Learning

$$GS_i = \alpha + \beta_0 GF_{ik} + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 \Omega^{fall}{}_i + \beta_5 \Omega^{spring}{}_i + \nu_i \qquad (4)$$

where:

GS_i : Grade earned by student *i* in Spring course who was taught by instructor *j* in the Fall.

GF_i: Grade earned by student *i* in Fall course taught by instructor *k.*

X_{1i} : demographic characteristics of student *i*.

X_{2i} : background academic characteristics of student *i*.

X_{3i} : personality characteristics of student *i*.

$\Omega_i^{fall} = \sum_{-ij=1}^{k-1} \varepsilon_{-\mathrm{ij}}$ is an instructor effect generated from first stage of the regression (done for spring also using the same method).

This regression yields valuable information in the interaction term between the personality intensity indices and the mean residual of the instructor, which gives a measure of the importance of personality traits given the fact that a student is randomly assigned a professor with a particular average residual, let's say Professor A. The interpretation of these values are similar to the personality indices; given a student has Professor A, a one unit increase in their personality intensity raises their GPA by the ɣ coefficient (in addition to the effect considered by the intensity index and average residuals alone). The intuition behind this is that if a student has a particularly harsh or easy instructor, we want to know how much of an additional role their personality will play. We see the effect of Professor A in his students even after he is no longer the student's instructor. The significance of the fall instructor residuals on the spring grade show that instructor standards matter, and that they matter in the context of sequential learning and higher education in general.

This result points to the relevance of soft skills in education and why soft skills for both students and instructors should be taken into account. It is not only the student's personality and how that impacts their success just as it is not only a teacher's tendencies to be a harsh or easy grader, but how these two interact with each other in the classroom and how this effects student performance over time.

IV: Data

The data set for this research contains extensive data on student outcomes and background characteristics, faculty information, and personality data. The student data set consists of roughly 41,000 students who attended the Naval Academy from 1988 to 2020[15]. The data we use considers both hard skills and soft skills as determinants of student success. Specifically, we include background academic and demographic information of students, personality tests of Midshipmen, and faculty data such as gender, whether they are civilian or military (and their accompanying rank), courses taught, and whether they are tenure or tenure-track. Outcome data for student success includes final grades in six sets of courses, GPA over freshmen spring and fall semesters, and overall GPA and class rank.

A. Student Data

The student data set contains extensive background information, including gender, minority status, whether the student was a recruited athlete, their weighted high school rank, SAT score, and their feeder source[16].

Below is a set of summary statistics for basic background characteristics of the data set:

[15] While there are ~41,000 student observations, the sample size for a given course ranges from ~7,000 to ~2,000 based on constraints of information for instructors and students and the ability to attribute residuals to instructors and the fact that students validate some classes.

[16] Feeder sources into the Naval Academy include NAPS (Naval Academy Preparatory School), foundation schools, boost (a program for enlisted sailors to commission as officers), nuclear (a program for prior-enlisted nuclear trained sailors to commission as officers

Table 1: Summary Statistics for Relevant Background Characteristics

	Observations	Mean	Std. Dev	Min	Max
Female	40911	.169	.375	0	1
Recruited Athlete	40911	.265	.441	0	1
Minority	40911	.231	.422	0	1
SAT Verbal	40911	642.321	73.046	230	805
SAT Math	40911	662.821	67.577	400	805
NAPS	40911	.167	.373	0	1
Foundation	40911	.059	.235	0	1
Boost	40911	.005	.070	0	1
Nuclear	40911	.009	.094	0	1
Weighted class rank	40911	560.435	132.006	137[17]	800

Some interesting takeaways from the background data of students are the mean values for Female (0.169) and Minority (.231), which help define our sample and the bias it may have. This means that approximately 16.9% of our sample population are female and 23.1% are considered a minority. This is where the sample of students at the Naval Academy may differ from that of a typical liberal arts institution—it is still a predominantly male, white population though it does increase in diversity each year with the class of 2019 graduating with a near 25% female class. The typical liberal arts institution has is 62.1% female and 53% white[18]. It is also interesting to

[17] These numbers are based off an algorithm that the Naval Academy uses to consider class rank across various sizes of high school.
[18] Data taken from the Department of Education (National Center for Education Statistics) and based off of top 5 liberal arts colleges that graduate the most students.

note that the average SAT Math score is 20 points higher than the average SAT Verbal score, which may indicate a natural aptitude for mathematics rather than humanities in our sample. For comparison, the average SAT scores at WestPoint are SATM=654, SATV=629[19]. The average SAT scores for the Ivy Leagues are SATM~792, SATV~736[20]. It is interesting to note that all three of these sets of data find that the SATM score is higher than the SATV regardless of school[21].

The academic performance parameter in this project is the student grade achieved in each class. The Naval Academy grades on a five-letter grading scale (A, B, C, D, F); there are no qualifiers such as "+" or "-" and final grades do not reflect percentage points (a student who earns an 80.0% will have the same GPA as a student who earns an 89.9%). These letter grades are coded in the data according to their associated GPA score. For example, an A is coded as a 4.0, a B is coded as a 3.0, and so on. A table of summary outcome statistics is shown below:

Table 2: Student GPA Summary Statistics

	Obs	Mean	Std. Deviation
Calculus I	7734	2.746	0.8953
English I	17949	2.950	0.7322
Chemistry I	20612	2.429	0.9129
Physics I	17754	2.755	0.9315
Government	20863	2.995	0.8130
History	22559	2.987	0.8206

[19] "Prep Scholar ." West Point SAT Scores and GPA, 2018, www.prepscholar.com/sat/s/colleges/West-Point-SAT-scores-GPA.
[20] Staffaroni, Laura. "Ivy League SAT Averages." Good SAT Scores: Ivy League Plus Edition, blog.prepscholar.com/good-sat-scores-ivy-league-plus.
[21] This phenomenon has been noted on a national level. As noted by Richard Rothstein in 2002, the U.S. especially seems to be widening a gap between the average SAT math and verbal score. Some hypotheses have been provided to explain the gap but there have been no definitive discoveries to explain the phenomena.

The class most often validated in Calculus I, which is why the number of observations is much lower, and the class least often validated is History, which is why the number of observations is much higher[22]. The range between indicates how many students validate a specific course as well as the loss of observations because we cannot identify a specific faculty member[23]. Humanities tend to have higher grades associated with them (all three nearly at 3.0, which would be a "B"), whereas the STEM course average grades tend to be lower, with the lowest average grade in Chemistry.

Because we want to focus on learning in sequential courses, this analysis follows six sets of courses: Calculus I and II, English I and II, Chemistry I and II, Physics I and II, and Government and History. In a typical student matrix, freshmen will take the first sequence of courses during their fall semester and the follow-on course in their spring semester (i.e. a student will take Calculus I, English I, Chemistry I, and Government in the fall and Calculus II, English II, Chemistry II, and History in the spring). The exception is physics; Midshipmen take physics in their second year following the same pattern (Physics I in the fall, Physics II in the spring). We follow these courses because of the sequential nature of their curriculum and because it represents a range of classes that students take and allows for a broader analysis than a single subject. Our selection of these courses also gives a range of what we consider more or less sequentially dependent. For example, Calculus I and Calculus II are more sequential in nature than English I and English II, so it should serve as a robustness check to compare, especially across STEM and humanities courses.

[22] All student have the opportunity to validate a class over a series of tests the summer before freshman year starts. Validation ensures a student has a good enough understanding of the subject material that it would not benefit them to take the course again. If a student validates Calculus I, they will be enrolled in Calculus II.
[23] There are some instructors listed only as "staff", so there is not a unique instructor code we can tie to them or their respective residuals.

There is a random assignment of students to each class and a random assignment of teachers so there is no opportunity for selection bias on the part of students because they cannot choose their section or professor. The algorithm used by USNA to create random sections (accounting for ethnicity, gender, SAT math and SAT verbal scores) will be checked using resampling methods to ensure accurate measurement. The assignment of instructors to random students is what allows for the use of average instructor residual in the analysis. The hypothesis is that, because students and instructors are randomly assigned, the average residual should capture the instructor effect. More specifically, a particular instructor will have underperformers, over performers, and average students in their class, thus the residual calculated across all students for particular professor should capture the grade distortion on average even accounting for the differing level of students.

B. Personality Data

For this project, we use personality traits (and specific characteristics of an overall personality trait) as a proxy for soft skills. The most significant addition of this research to other research in this field is the inclusion of personality as a factor in student success across an entire population[24]. Our data set includes the result of each student's MBTI (Myers-Briggs) personality test. Every freshman takes the Myers-Briggs Personality test and records their results for Leadership class (there are approximately 40,000 observations for this data).

The Myers-Briggs personality test has typically been the standard for personality tests used in educational research, especially for higher education. The Myers-Briggs Test, referred to as MBTI, was designed by Isabelle and Katharine Briggs based on Carl Jung's personality theory

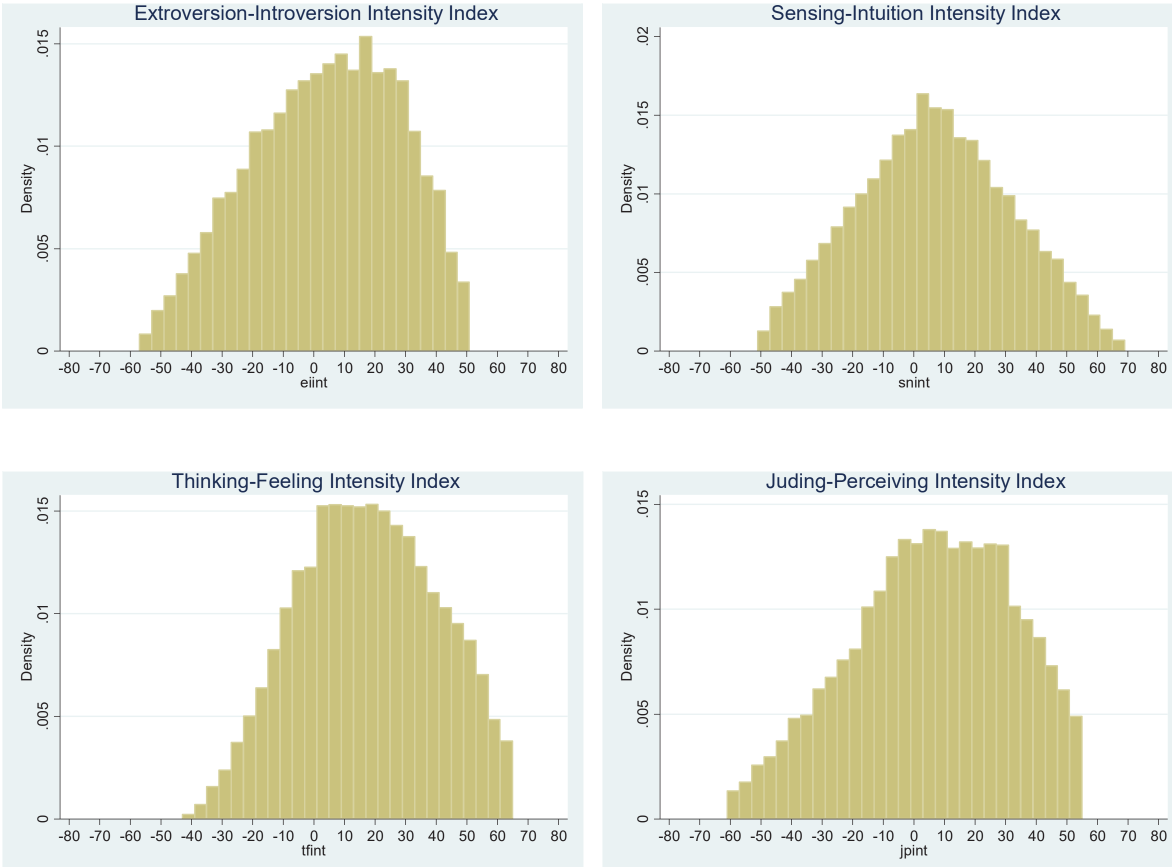
[24] This builds upon the previous research mentioned, specifically Borg and Shapiro (1996) and Emerson and Taylor (2007), by including more than a single class in the MBTI study and looking at the effects across six courses over two semesters

and was designed to make the insights of his theory applicable to individuals and groups to understand their natural preferences and thus create better teams and more understanding communities. The MBTI test is a series of questions that ultimately places one on a scale of four dichotomous personality preferences. The four areas that the test assesses one's preferences are: where people draw their energy (Extraversion or Introversion); how people draw in information (Sensing or Intuition); how people make decisions (Thinking or Feeling); and how people like to structure their outer life (Judging or Perceiving). While people may use both sides of a dichotomous pair, each person has inherent preferences that the test highlights. A test taker is scored from 0 to100 on each category in their dominant domain, with 0 being a weak preference for one trait versus the other and 100 being the strongest preference. There are 16 different potential personality types from these four dichotomous options. A summary table of the four pairs is provided in the appendix.

For each personality distinction that a test taker receives, they have a matching intensity index. For example, if a student's personality test is ESTJ, they will have four corresponding numbers with the letters (E/I=38, S/N=17, T/F=50, J/P=22). The numbers range from 0-100, with 0 being the closest to neutral between the two dichotomous options and 100 being the most intensely one or the other.

Each measure is mapped to a scale of (-100, 100), and is shown below:

Figure 1: MBTI Personality Intensity Indices



The x-axis on these graphs is the intensity index for each respective set of traits. The range of values is (-100, 100), but the graphs above range from (-80,80) to more easily view the distribution. The y-axis is the frequency that a specific score appeared. While all four personality traits appear to be normally distributed, there is a range of approximately 110 for each index indicating that there is no trait that is completely dominant in the population. A table of summary statistics is shown below:

Table 3: MBTI Characteristics Intensity Indices

MBTI Characteristics Intensity Indices					
	Observations	Mean	Std. Deviation	Min	Max
Extraversion-Introversion	36768	2.90	24.14	-57	51
Sensing-Intuition	36768	5.78	25.00	-51	67
Thinking-Feeling	36768	16.55	22.66	-43	65
Judging-Perceiving	36768	4.84	26.20	-61	55

These statistics show that Midshipmen tend towards Extraversion, Sensing, Thinking, and Judging (personality type ESTJ), as seen by all positive average values. The most heavily skewed of the four traits is in the Thinking-Feeling index with the highest mean value. It is interesting to note that military officers tend towards –STJ (either introverted or extroverted); over 66% of military officers tend towards that whereas only about 20% of the general adult population represents ESTJ and ISTJ combined[25]. Despite this difference from the general population, the data set is large (~40,000 observations) and contains large enough variance that any conclusions are likely to have significance for the general population. We also see that, on average, while our population tends towards ESTJ, the mean values for all pairs except Thinking-Feeling all hover near zero, so the data will still exhibit external validity.

C. Faculty Data

The faculty data set contains approximately 3200 observations of instructors who have taught at the Naval Academy. The data set includes their current job area, current job title, their gender, whether an instructor is civilian or military (and their accompanying rank and branch if they are)

[25] MBTI Basics." The Myers & Briggs Foundation , 2018, www.myersbriggs.org/my-mbti-personality-type/mbti-basics/home.htm?bhcp=1.

and whether the instructor is tenure track. Because of the extensive data set we can consider analysis on the effect of gender and having a military instructor.

A summary table is shown below.

Table 4: Instructor Summary Statistics: Percentage of Female and Military Instructors

	Female	Military
Overall	18.6	59.8
Calculus I	17.8	45.4
English I	37.8	35.7
Chemistry I	35.1	26.6
Physics I	14.4	34.7
Government	19.9	62.1
History	13.1	55.7

Notes: Overall includes only the courses in this study, not all USNA faculty.

From this table, it is interesting to note that certain subjects tend to have more female instructors or military instructors. The larger number of military instructors in Calculus I, Government, and History is likely due to the fact that junior officers serve as rotational instructors on their shore tours, often here for less than three years, after which time they may again be replaced by a short-term instructor.

V: Results

Through our analysis, we find the soft skills of students and the soft standards of teachers matter for student success in higher education. Our results show that certain personality traits can help explain student achievement, specifically the impact of being a “judger”. We also find that teachers have significant effects in this model. Excessively “cushy” instructors hurt student performance in follow-on courses. Results also show that certain personality types may help students overcome excessively easy instructors and that the gender of the student and instructor play a role in the ability of a student to succeed in a follow-on course regardless of the standards of their initial instructor.

A. The Effect of Soft Skills On Overall Student Performance

Before looking at the impacts of student personality with instructor quality, we look at the effects of student personality on student success more holistically. In order to see the full effects of personality on Midshipmen academic performance, we analyze multiple educational outcomes using a variety of hard and soft skill measures in order to effectively measure educational outcomes in different ways.

The primary educational outcomes are the six initial courses that we analyze throughout the project (Calculus I, English I, Chemistry I, Physics I, Government, History) as well as four more comprehensive measures: overall fall academic GPA freshman year, overall spring academic GPA freshman year, class rank at time of graduation, and GPA at the time of graduation.

In the first specification (1), we include only the base background characteristics that have been the most significant predictors of academic achievements (math SAT score, verbal

SAT score, a weighted high school rank, and an indicator variable accounting for a student's gender and minority status as well as their status as a recruited varsity athlete). The second specification includes these background characteristics as well as the personality intensity indices. The output table for the overall GPA is shown below:

Table 5: Personality Effect on Final GPA

	Dependent Variable: Final GPA	
	(1)	(2)
Female	0.0217**	0.00739
	(0.00893)	(0.00899)
Minority	-0.0863***	-0.0902***
	(0.00809)	(0.00805)
Recruited Athlete	-0.0327***	-0.0386***
	(0.00789)	(0.00791)
SAT math	0.00259***	0.00254***
	(0.0000585)	(0.0000583)
SAT Verbal	0.000872***	0.00104***
	(0.0000536)	(0.0000542)
High School Rank	0.000907***	0.000886***
	(0.0000268)	(0.0000268)
EI Intensity Index		0.000402***
		(0.000139)
SN Intensity Index		0.000854***
		(0.000147)
TF Intensity Index		-0.0000806
		(0.000156)
JP Intensity Index		0.00213***
		(0.000139)
_cons	0.0661*	0.0162
	(0.0391)	(0.0394)
N	39194	35432
R-sq	0.174	0.207

Table Notes: Standard Error in parentheses *p<.1 **p<.05 ***p<.01

Table 5 (overall GPA at graduation) offers the most comprehensive view of academic achievement over students' time at USNA[26]. Model 1 shows student background characteristics

[26] The rest of the output tables from these regressions are available in the appendix and their results are similar

on student success that only takes into account "hard skills". Model 2 accounts for these in addition to "soft skills" as proxied by the MBTI personality intensity indices. It is important to note that when including the proxied soft skills into the regression, the values for the typical hard skills stay nearly the same as they were in Model 1 despite the fact that these indices appear statistically significant on the p=0.01 level. A change that does stand out between Model 1 and Model 2 is the effect of gender. From Model 1 to Model 2, the value of Female, while remaining positive, reduces from 0.0217 to 0.00739 and changes from being statistically significant at the p=0.05 level to dropping below the threshold of p=0.1. This change in coefficients indicate that in Model 1, a female is predicted to achieve a grade 2% higher than her male counterpart all else being equal; the change when personality traits are taken into account drops this expected value to 0.7% higher grade than her male counterparts. This could indicate that some of the effects of being female that boosted academic performance are captured in the personality measures. This observation is in line with the previous research[27]. This may mean that females are more likely to have specific personality traits than their male counterparts, which are affecting their performance more than solely their gender.

Model 2 offers us our initial insight into the importance of personality traits for student achievement. Three of the four personality intensity indices (Extraversion-Introversion, Sensing-Intuition, Judging-Perceiving) are statistically significant at the p=0.01 level. Most noticeable is the value of "jpint", which is the intensity index of being a "judger" versus a "perceiver" (a student's their preferred level of structure), which appears statistically significant across all educational outcome measures we tested including the six individual courses, freshman fall GPA, freshman spring GPA, and overall AOM (included in appendix). We also see that certain

[27] Borg and Shapiro (1996)

traits of a dichotomous set appear to lead to higher student achievement. All positive coefficients show that the first listed personality trait is more helpful for academic achievement whereas negative coefficients mean that the second listed personality trait aids student achievement more (this is by nature of the way the intensity indices are coded). Based on this finding, "extraversion" helps more than "introversion", "sensing" helps more than "intuition", "feeling" helps more than "thinking", and "judging" helps more than "perceiving".

These regressions are especially useful because they are done across all student observations (with n=35432), giving a very robust data set that allows us to see major trends across two decades of data. Based on these values we can see that personality does influence student performance. This takeaway points to the fact that soft skills should be considered alongside hard skills by institutions of higher education in order to predict success or help students along the way. Even if these soft skills are not used as a metric to determine placement, it would still be useful for a student to know where he or she may struggle or excel more naturally as they look to play on their strengths. This result is consistent with previous research that indicates that student personality and soft skills do contribute to student success. This result is helpful in its comparison of the two models listed above (one with and one without soft skills), and serves to add an additional factor to the previous research, showing that the inclusion of soft skills contributes to the overall robustness of the function of academic success.

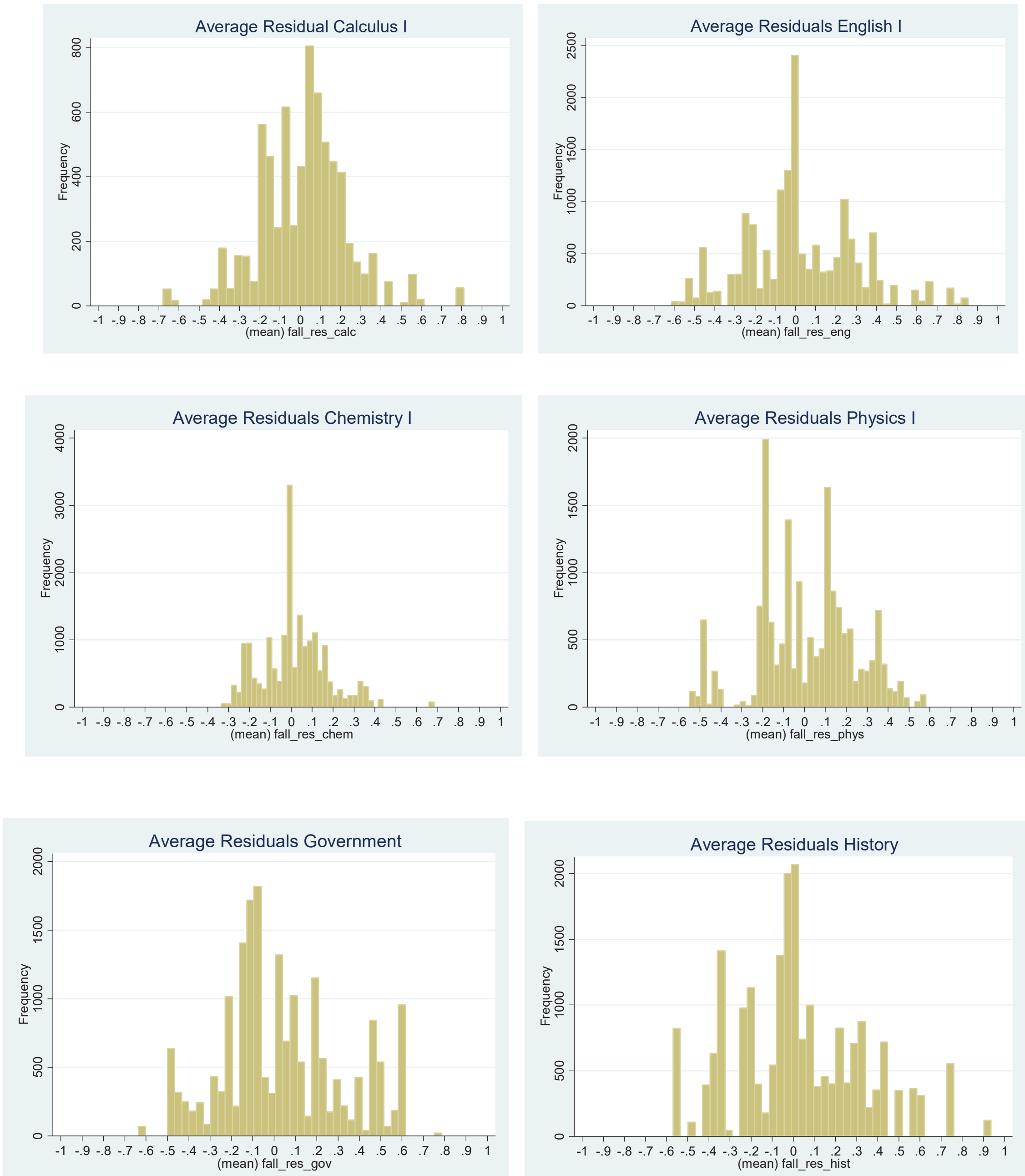
After determining the importance of student personality on a larger scale, we look more closely at student performance through our six sets of sequential courses: Calculus, English, Chemistry, Physics, Government, and History.

B. Soft vs Hard Standards Index by Faculty in Sequential Learning

We now look to the effects of personality and how personality traits interact with faculty standards across different courses. For each course and student, we use both soft and hard skills to predict what their grade in a specific course will be. After finding the difference in expected grades and awarded grades, we are able to aggregate these differences across all students that an instructor has taught, and thus find the average residual for each instructor. This residual is what we consider the grade distortion of each professor. It is unique to the instructor based on his or her own standards. The histograms below show these values by course so that we can see a distribution of grade distortion among different courses and see if any trends appear.

The graphs below show a breakdown of average residuals by course.

Figure 2: Average Residuals by Course



The x-axis of these graphs is the mean residual by specific instructor. The y-axis is the frequency of the residual. For all six courses, the residuals center on zero but there is a lot of variation.

SM121 has the largest range of residual values, with values down to -0.7 and values above 0.8[28], which represents over half of a letter grade in both directions. Below is a table of the summary statistics for the instructor effects:

Table 6: Instructor Residual Summary Statistics

Course	Observations	Mean	Std. Dev	Min	Max
Calculus I	473	.007	.246	-.679	.811
English I	132	.069	.315	-.614	.854
Chemistry I	99	.026	.183	-.335	.683
Physics I	117	.024	.244	-.549	.582
Government	110	.034	.279	-.636	.786
History	93	.088	.301	-.569	.938

Note: These are calculated from Equation (3) and are the fall average residuals across all instructors in a given course. A t-test was conducted to see if the average residuals for each course were statistically different from zero. The only classes that are statistically different than zero are English and Government.

First we see that average residuals hover close to zero across all courses, which is exactly what we would expect them to do as an average across all professors and students (meaning that, on average, students tend to achieve the grade in the class that they are predicted to achieve based on background characteristics and personality type). We ran a t-test to test if the average instructor effect in each class are statistically different than zero; we find that the only classes that are statistically different than zero are English and Government. Next, Calculus I stands out as having a much larger number of observations than the other courses. This is due to the high

[28] This is significant as it is nearly a full letter grade numerically, and because of USNA's grading system of only full letter grades, this would raise or lower their GPA in that class by either +1.0 or -1.0

number of military professors[29] that teach this course. Lastly, we see that Calculus I has the largest range of residuals and that Chemistry I has the smallest. We also see that humanities courses (English, History, Government) have the biggest standard deviation values, potentially because grading is more subjective in these courses than their STEM counterparts.

It is also valuable for research in this institution to distinguish between civilian and military professors. Below are two tables that differentiate civilian versus military by course:

Table 7: Standards Index of Civilian vs Military Instructors

	Civilian		**Military**	
	Observations	Mean Residual	Observations	Mean Residual
Calculus I	197	0.004	276	0.009
English I	48	0.079	84	0.064
Chemistry I	44	0.012	55	0.037
Physics I	49	0.027	68	0.022
Government	50	0.019	60	0.047
History	40	0.058	53	0.111

Note: This table encapsulates the same data presented in Table 6, but it is split between civilian and military instructors in order to see if there appear to be different standards across the groups. A t-test was conducted to see if the average residuals differ across these two groups, which they do not.

From these tables we see that there are more military than civilian professors across all courses. This, again, is due to the introductory nature of all of these courses and the high rate of turnover for military professors. Through conducting a t-test, we find that the differences in these two groups (civilian versus military) is not statistically significant in any course, although this could potentially be due in part to the limited number of observations. The summary statistics and

[29] SM121 is an introductory level class that most officers are qualified to teach, so these are the junior officers that rotate every two to three years.

dispersion of average residuals across courses analysis for the follow-on course.

C. Soft Skills and Soft Standards: Interactions in the Classroom and Learning Outcomes

After a course-by-course analysis using the methods outlined in Equations (2) and (3), we confirm the impact of background characteristics and hard skills (those that are typically taken into account when determining student outcomes) as well as the importance of soft skills in students and the impact of soft standards from instructors. The output tables of the second stage regression, Equation (4), are shown below. For ease of reading, they are split into two tables from the same regression. The first table includes all of the background traits from the regression while the second table includes the average residual, the fall grade, and all personality information.

Table 8: Background Characteristics of Soft Skills and Soft Standards Second Stage

	Spring Courses GPA					
	Calculus II	English II	Chemistry II	Physics II	History	Government
SAT Math	0.000870***	0.000284***	0.00132***	0.00168***	-0.0000289	0.000664***
	(0.000203)	(0.0000903)	(0.0000889)	(0.000106)	(0.000120)	(0.000118)
SAT Verbal	-0.000599***	0.00131***	0.0000894	-0.000117	0.00154***	0.00127***
	(0.000144)	(0.0000913)	(0.0000752)	(0.0000881)	(0.000117)	(0.000114)
Weighted HS Rank	0.000566***	0.000586***	0.000464***	0.000707***	0.000447***	0.000628***
	(0.0000713)	(0.0000406)	(0.0000376)	(0.0000446)	(0.0000549)	(0.0000527)
Female	-0.0306	0.0980***	-0.0460***	0.00472	-0.187***	0.0645***
	(0.0238)	(0.0135)	(0.0120)	(0.0142)	(0.0178)	(0.0176)
Minority	-0.0326	-0.100***	-0.0405***	-0.0454***	-0.0845***	-0.0952***
	(0.0216)	(0.0122)	(0.0110)	(0.0132)	(0.0161)	(0.0160)
NAPS	-0.145***	-0.0147	-0.0710***	0.0365**	-0.0324	-0.0734***
	(0.0278)	(0.0162)	(0.0153)	(0.0179)	(0.0213)	(0.0209)
Foundation	-0.159***	0.0200	-0.150***	-0.0487**	-0.00240	-0.0589**
	(0.0402)	(0.0223)	(0.0210)	(0.0238)	(0.0303)	(0.0288)
BOOST	-0.378	-0.0506	0.232	0.360	0.0492	0.474
	(0.499)	(0.213)	(0.222)	(0.327)	(0.307)	(0.430)
Nuclear	0.142**	0.0299	0.220***	0.209***	0.00427	0.0277
	(0.0607)	(0.0391)	(0.0364)	(0.0478)	(0.0537)	(0.0511)
Recruited Athlete	-0.0365*	-0.0445***	-0.0278**	-0.0500***	-0.0671***	-0.0444***
	(0.0213)	(0.0121)	(0.0109)	(0.0126)	(0.0160)	(0.0156)

Table Notes: Standard Error in parentheses *p<.1 **p<.05 ***p<.01
Dependent Variable: Final Spring Grade in Calculus, English, Chemistry, Physics, History, and Government
Independent Variables: SAT (Math and Verbal), high school rank, gender, recruited athlete status, minority status, and a variable for their entrance location (NAPS, foundation, BOOST, or nuclear).

From this table, background characteristics that initially appear important are SAT scores, high school rank (adjusted for size of school), and whether the students is a recruited athlete or a minority. We see that a student's SAT math score is statistically significant at the p=0.01 level and positive across all courses except government, where is it negative and not statistically significant. We see a similar trend for SAT verbal scores: they are positive and statistically significant across all courses except Physics (where it is negative and not statistically significant), and Calculus, where it is negative and significant at the p=0.01 level. The SAT scores have been standardized and analyzed using beta coefficients, where we can see the importance of SAT math versus SAT verbal in STEM vs non-STEM classes[30]. For example, the beta coefficient in Calculus for SAT Math is 0.05 whereas the coefficient for SAT verbal is -0.05, so we see that in Calculus, a higher SAT Math score is beneficial, but a higher SAT Verbal score is detrimental to a student's grade in Calculus.

We see that being a female is a detriment to academic success in Calculus (though not statistically significant), Chemistry, and History. The most consistent demographic across all courses is the effect of being a recruited varsity athlete: it is negative and statistically significant across all courses. These confirm what most research typically uses as their predictors of student success, that these demographic and academic background values matter, but we further analyze and find that soft skills also account for a portion of student success.

The following table shows the output from the regression that is centered on the soft skills of students and teacher standards, including the effect of the fall grade on spring grade, the effect of the student's fall grade instructor's grading tendencies, and the student's soft skills.

[30] Beta coefficients are used as a way to standardize the variables and thus give us the ability to make a more direct comparison between the coefficients associated with dependent variables.

Table 9: Soft Skills and Soft Standards Regression Output Second Stage

	Spring Courses GPA					
	Calculus II	English II	Chemistry II	Physics II	History	Government
Final Fall GPA	0.576***	0.395***	0.684***	0.540***	0.397***	0.385***
	(0.0118)	(0.00817)	(0.00601)	(0.00723)	(0.00998)	(0.00970)
Fall Instructor Effect	-0.561***	-0.356***	-0.433***	-0.461***	-0.397***	-0.360***
	(0.0435)	(0.0206)	(0.0301)	(0.0244)	(0.0262)	(0.0239)
Spring Instructor Effect	1.118***	0.954***	0.810***	0.957***	0.966***	1.065***
	(0.0400)	(0.0195)	(0.0283)	(0.0236)	(0.0231)	(0.0235)
EI Intensity Index	-0.0000568	0.000451**	-0.000383**	-0.000270	0.000121	0.000458*
	(0.000393)	(0.000212)	(0.000195)	(0.000225)	(0.000284)	(0.000275)
SN Intensity Index	0.000622	-0.000921***	-0.000279	-0.000520**	0.000442	-0.000580**
	(0.000420)	(0.000226)	(0.000208)	(0.000237)	(0.000301)	(0.000293)
TF Intensity Index	-0.000236	-0.000329	-0.0000618	-0.000146	-0.000425	0.000416
	(0.000445)	(0.000240)	(0.000219)	(0.000254)	(0.000322)	(0.000309)
JP Intensity Index	0.00174***	0.00199***	0.00132***	0.00172***	0.00235***	0.00179***
	(0.000394)	(0.000215)	(0.000197)	(0.000227)	(0.000288)	(0.000279)
_cons	0.552***	0.509***	-0.563***	-0.222***	0.660***	0.293***
	(0.145)	(0.0738)	(0.0647)	(0.0758)	(0.0942)	(0.0920)
N	6093	14792	16707	15103	8561	8968
R-sq	0.414	0.368	0.597	0.485	0.409	0.420

Table Notes: This analysis takes into account year effects as well as a chi-squared test between courses.
Dependent Variable: Final Spring Grade in Calculus, English, Chemistry, Physics, History, and Government
Independent Variables: Fall final grade, fall instructor average residual (instructor effect), personality intensity indices.

Analysis of Table 9 shows the importance of the first course final grade, which is not surprising; if a student does particularly well in Calculus I, we also expect them to perform well in Calculus II. We see that this is statistically significant at the p=.01 level across all courses with a maximum value of 0.684 in Chemistry and a minimum value of 0.385 in Government. This means that if a student receives one letter grade higher in Chemistry I, their grade in Chemistry II will rise 0.684 on average out of a 4.0 scale, which is nearly two thirds of a letter grade increase. Noteworthy in this findings is that the fall grade affects the spring grade more in all STEM classes than it does in humanities classes, which is what we might expect given the sequential nature of the STEM topics. We can see how it makes sense that a student who was pushed harder to understand basic derivatives in Calculus I will benefit from that work in Calculus II as the curriculum builds upon that knowledge. By contrast, the harshness of a particular instructor in English I may have helped develop a student's writing to improve and thus helped them in English II, it is not the same sort of sequential building that we in the STEM courses where the student must understand previous material before understanding further.

Most striking from the table, however, are the coefficients associated with the average residual of the professor. These values for fall instructor effect are negative for all six sets of courses, ranging from -0.356 in English (the grade least impacted by a first-course instructor) to -0.561 in Calculus (the grade most impacted by a first-course instructor). This means that if an instructor has an average residual of 0.5, a "B" (80%) student could expect to get a higher "B" (86%) in Calculus I, but in the follow on course could expect a "C" (77%).

Given a student has a professor who is a "cushy" grader (or is arbitrarily inflating grades or teaching to the test), they are doing their students a disservice in the following semester. If we break down the courses into STEM vs humanities, we see once again that STEM courses are more heavily impacted, with Calculus, Chemistry and Physics having a higher absolute value than English, History, and Government. Calculus is most impacted at a value of -0.561, indicating that a student will suffer a loss of nearly half of a letter grade in the follow-on course[31].

These results do not necessarily indicate that the teacher is not teaching as effectively, it could mean they are teaching to the test or that the teacher tends to round up their grades more or has softer standards. This tendency then means that the student receives the signal that they do not need to put in as much effort to receive a better grade and that decreased effort level may carry over into the following course. If, for example, a student who was expecting to get a C in Calculus I based on the amount of effort they were putting into the class, ends up with a B, they now associate their C-level of work with a B-level outcome and adjust their expectations accordingly. The next semester, this student likely chooses to continue to put forth the same amount of effort because they are satisfied with the result from last semester. However, the student may not truly understand all the material of Calculus I because they did not have an incentive to put in necessary work to understand concepts with which they may have struggled. This potential lack of understanding has now followed this student into their follow-on course based upon the signal from the instructor that their initial C-level work warranted a B-level outcome. In this context, soft standards of teachers harm their student's academic performance in follow-on courses.

For the spring instructor effect, we see that the effect is positive and statistically significant at the p=0.01 level across all courses. The values range from 0.810 in Chemistry II

[31] This is given off of the indication that an instructor has a 1.0 fall average residual.

(the grade least impacted by second-course instructor) to 1.118 in Calculus II (the grade most impacted by second-course instructor).[32]

Using beta coefficients, we see that spring instructors in STEM courses tend to have approximately twice the impact of fall instructors in the same course. For example, in Calculus, the beta coefficient for fall instructor effect is -0.13 whereas the beta coefficient for spring instructor is 0.27. Likewise in Physics, the beta coefficients for fall and spring instructors are -0.12 and 0.24, respectively. In non-STEM courses, the spring instructors tend to have just over twice the impact of the fall instructors. For example, in English, the beta coefficients for fall and spring are -0.13 and 0.33, respectively and in History the coefficients are -0.14 and 0.36, respectively. These coefficients point out that while the fall instructor has a definitive effect, the impact is not as significant on student performance in the spring course as the impact of the spring instructor. This result makes sense: the immediate effect of a "cushy" spring teacher in the spring course will outweigh the negative effect of a "cushy" fall instructor, though both will play a role. [33]

[32] It is interesting to note that Calculus is the most heavily impacted by instructor standards, but on average we tend to see that spring instructor is approximately twice as impactful as fall instructor, so it would make sense that because it has the greatest magnitude for fall instructor, it would also show the largest impact for spring instructor.

[33] We also ran this regression without accounting for spring instructor effect, and we see that the inclusion of this variable does not noticeable change the other coefficients (especially the impacts of fall final grade and fall instructor effect).

D. Gender and outcomes

After seeing the effect of student soft skills, professor tendencies, and the interaction of the two, we look more carefully at the role of instructor and student gender in student grade outcomes. Based on previous research, we expect that gender may play some role in student success in certain courses, particularly in STEM courses. For these regressions, we use the same initial regressions as in the previous section but limit the sample based on the gender of the student and the instructor. The four categories tested are female student-female instructor, male student-male instructor, female student-male instructor, and male student-female instructor. We test the effect of gender across all courses overall and split for STEM and non-STEM courses.[34] This analysis allows us to look at our results in the context of another paper written about the effect of gender (particularly of the instructor) in STEM courses. Carrell et al. (2009) find that female students are more likely to pursue STEM majors if their initial teacher in a STEM course is a female because it offers them a role model in the field[35]. Our results offer a perspective that, although they may be more likely to join the field, female students with female instructors are the group most susceptible to suffering in a follow-on course due to soft standards of instructors.

The table below shows the effect of gender across all courses.

[34] We also test on a course-by-course basis, but because the sample of female students and female instructors is smaller, it is difficult to get a statistically significant result for an individual course. The tables for Calculus and English are included in the appendix.

[35] Sex and Science: How Professor Gender Perpetuates the Gender Gap Scott E. Carrell, Marianne E. Page, and James E. West NBER Working Paper No. 14959 May 2009 JEL No. I20,J24

Table 10: Overall GPA

	Overall GPA			
	Spring GPA			
	Female-Female	Male-Male	Female-Male	Male-Female
Fall Final GPA	0.656***	0.568***	0.573***	0.615***
	(0.0138)	(0.00406)	(0.00845)	(0.00716)
Average Fall Residual	-0.345***	-0.319***	-0.302***	-0.245***
	(0.0536)	(0.0161)	(0.0305)	(0.0302)
EI Intensity Index	-0.00104**	0.0000481	0.0000466	-0.000108
	(0.000493)	(0.000126)	(0.000269)	(0.000233)
SN Intensity Index	0.000243	0.0000130	0.000181	-0.0000192
	(0.000548)	(0.000132)	(0.000295)	(0.000247)
TF Intensity Index	-0.000147	-0.000119	-0.000388	-0.000420
	(0.000542)	(0.000143)	(0.000296)	(0.000268)
JP Intensity Index	0.000859	0.00125***	0.00159***	0.00112***
	(0.000526)	(0.000125)	(0.000288)	(0.000233)
EI Intensity*Fall Residual	0.000181	-0.00159***	0.000274	-0.00150*
	(0.00186)	(0.000489)	(0.00105)	(0.000906)
SN Intensity*Fall Residual	-0.0000974	-0.00112**	-0.000122	0.00120
	(0.00212)	(0.000514)	(0.00114)	(0.000967)
TF Intensity*Fall Residual	-0.00300	0.00152***	-0.000144	-0.00291***
	(0.00215)	(0.000558)	(0.00116)	(0.00106)
JP Intensity*Fall Residual	0.00223	0.00131***	0.000548	0.00169*
	(0.00199)	(0.000484)	(0.00113)	(0.000908)
_cons	0.0141	-0.0298	0.0259	-0.103
	(0.153)	(0.0426)	(0.0891)	(0.0745)
N	4306	55382	12946	17455
R-sq	0.431	0.384	0.370	0.414

Table Notes: Dependent variable: Overall GPA across the six second-semester courses studied. Independent Variables shown: final GPA in the fall (overall), average residual of fall instructors (overall), intensity indices of personality, and interaction terms of residuals and personality terms. Background characteristics were included in the regression, but not displayed for ease of viewing.
First gender listed is student gender, second gender listed is instructor gender.

From this table, we see that the fall grade is still the most significant predictor of spring grade; it is statistically significant and positive across all subgroups. The maximum magnitude is for the female-female subgroup (0.656) and the smallest is for the male-male subgroup (0.568). We also see that fall average residual is statistically significant and negative across all subgroups, mimicking the trend we saw before splitting by gender. The maximum magnitude is for the female-female subgroup (-0.345) and the minimum is for the male-female subgroup (-0.245). We see that the Judging-Perceiving Index remains the most important personality index for student success. It is statistically significant across all subgroups except female-female (which could be due to a more limited sample size), and is positive for each group. We also see that all four interaction terms are statistically significant, but only for the male-male subgroup, meaning that certain personality traits (Introversion, Intuition, Thinking, Judging) could help male students overcome a "cushy" male teacher.

To go into more detail, we split the sample between STEM and non-STEM courses to compare. The table below shows the output of STEM vs non-STEM courses.

Table 11: GPA for STEM and Non-STEM Courses

	Spring GPA							
	STEM				**NON-STEM**			
	FF	MM	FM	MF	FF	MM	FM	MF
Final Fall GPA	0.651***	0.635***	0.610***	0.662***	0.382***	0.398***	0.372***	0.414***
	(0.0193)	(0.00572)	(0.0119)	(0.00970)	(0.0210)	(0.00573)	(0.0124)	(0.0105)
Fall Average Residual	-0.577***	-0.334***	-0.398***	-0.477***	-0.125**	-0.147***	-0.0908***	-0.113***
	(0.0808)	(0.0233)	(0.0475)	(0.0395)	(0.0549)	(0.0142)	(0.0311)	(0.0269)
EI Intensity Index	-0.000760	-0.000185	-0.0000612	-0.000307	-0.000859	0.000489***	0.000264	0.000181
	(0.000647)	(0.000189)	(0.000386)	(0.000323)	(0.000609)	(0.000158)	(0.000346)	(0.000293)
SN Intensity Index	-0.000712	-0.000141	0.000296	-0.0000114	0.000693	0.000197	0.000119	0.000115
	(0.000724)	(0.000198)	(0.000420)	(0.000347)	(0.000680)	(0.000167)	(0.000382)	(0.000309)
TF Intensity Index	-0.000669	-0.000183	-0.000401	-0.000737**	-0.000106	-0.000117	-0.000313	-0.000706**
	(0.000713)	(0.000214)	(0.000424)	(0.000368)	(0.000667)	(0.000180)	(0.000383)	(0.000336)
JP Intensity Index	0.00242***	0.00138***	0.00184***	0.00140***	0.000823	0.00159***	0.00190***	0.00146***
	(0.000688)	(0.000189)	(0.000412)	(0.000322)	(0.000653)	(0.000158)	(0.000371)	(0.000295)
_cons	-0.696***	-0.256***	-0.172	-0.191*	0.790***	0.372***	0.307***	0.346***
	(0.213)	(0.0642)	(0.129)	(0.107)	(0.197)	(0.0541)	(0.115)	(0.0966)
N	1939	23306	5690	7683	2367	32076	7256	9772
R-sq	0.503	0.481	0.448	0.509	0.243	0.289	0.252	0.289

Table Notes: “F” =“Female”; “M” = “Male”; first gender listed is student, second gender listed is instructor.
Dependent Variable: Overall spring GPA split between STEM and non-STEM courses. For example, the STEM overall grade includes Calculus II, Physics II, and Chemistry II.
Independent variables shown: Fall overall GPA (split STEM/non-STEM), average residual, personality indices[36].

From this comparison table, we can see interesting trends across STEM and non-STEM courses.

[36] For this table, we do not include interaction terms in order to isolate the effects of personality and average residual so we can more accurately see the effect with respect to gender. In order to ensure this still allows for an accurate regression, we tested multiple regressions with and without the interaction terms to test whether the other values remained similar, which they did.

First, we see that the fall grade is more influential for student success in STEM courses, ranging from 0.610 to 0.662 whereas in non-STEM courses it ranges from 0.372 to 0.414. Next we see that the values for the average fall residual are larger in magnitude for STEM courses across all subgroups, ranging from -0.334 to -0.577 across STEM courses and only -0.0908 to -0.147 across non-STEM courses. These two combine to add weight to the narrative that the sequential nature of learning in STEM courses mean that both the fall grade and the standards of fall teacher both heavily influence student success in the spring.

We see that the female-female subgroup has the most negative value for fall average residual, meaning they are most affected by having a "cushy" teacher. We see this especially in the STEM courses. This means that female students who have a female instructor and receive a good grade in a STEM course (perhaps better than they deserve) are especially likely to suffer in their follow-on course. In the context of Sex and Science, (Carrell et al., 2009) female students may be incentivized to join STEM majors by the signals they receive from their female instructors (in the form of higher grades), but they suffer in later courses. This is an aspect of student performance the Sex and Science paper does not consider in their findings: the ability to test the success of these students in follow-on courses, and the role that gender plays in that success.

VI: Conclusion

Our findings show promising evidence that introductory professors significantly affect student performance. Specifically, we see that excessively "cushy" instructors hurt student performance in follow-on courses in both STEM and Humanities courses. While we find this result across all classes, we find that when we split the sample by gender that the negative impacts of a "cushy" teacher are primarily a male instructor-male student phenomenon.

Our findings also shows that student personalities impact student performance, but to a lesser degree than instructor effects. We see that personality types do account for aspects of student performance, especially when looking at cumulative measures. More specifically, we see that extroverts perform better in non-STEM courses and introverts perform better in STEM courses, "thinkers" outperform "feelers" in STEM courses, and "judgers" outperform "perceivers" across all courses. We see that soft skills tend to matter more in non-STEM courses than they do in STEM courses. We also see that certain personality types may help students overcome excessively cushy instructors. The most consistent finding is the significance of the judging-perceiving index, specifically the benefit of being a "judger": it is the only personality trait that is beneficial across all subjects, cumulative GPAs and overall class rank.

Looking forward, we are looking to build a model of teacher inputs and consider the tradeoffs between motivation and instruction, and the possibility that effectiveness of either approach depends on student characteristics. We are also looking to analyze the effects of having a civilian versus military instructor and how that, when coupled with gender, may affect academic success. We hope to correlate what we have already found about teacher characteristics with student evaluation forms and see what relationship appears, if any. We are also looking to consider the breadth of the instructor effects and other metrics to see a professor's value-add.

One way to do this is looking at the effects of teacher quality on major selection, both preliminary and final. We are also trying to see the longer-run impacts of first-year instruction on performance and consider other types of soft-skill measures, such as aptitude or OCEAN scores that are also present within the data.

This research points to the fact that soft skills should be considered in education. Our current national education policy, No Child Left Behind, focuses primarily on the development of hard skill in students and does not leave much leniency in curriculum opportunity for instructors. Both the soft skills of students and the standards that instructors enforce affect educational attainment in students, and should be brought into the conversation at lower levels of education.

VII: References:

Borg, M. O., P. M. Mason, and S. L. Shapiro. 1989. The case of effort variables in student performance. Journal of Economic Education 20 (Summer): 308-13.

Charkins, R. J., D. M. O'Toole, and J. N. Wetzel. 1985. Linking teacher and student learning Styles with student achievement and attitudes. Journal of Economic Education 16 (Spring): 111-20.

Carrell, Scott E., and James E. West. "Does professor quality matter? Evidence from random assignment of students to professors." Del Mar, Calif.: Prometheus Nemesis *Journal of Political Economy* 118.3 (2010): 409-432.

Carrell, Scott E., Page, Marianne E., and James E. West Sex and Science: How Professor Gender Perpetuates the Gender Gap NBER Working Paper No. 14959 May 2009 JEL No. I20,J24

Carrell, S. E., Sacerdote, B. I. and West, J. E. (2013), From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. Econometrica, 81: 855–882. doi:10.3982/ECTA10168

Carrell, Scott E., Teny Maghakian, and James E. West. 2011. "A's from Zzzz's? The Causal Effect of School Start Time on the Academic Achievement of Adolescents." *American Economic Journal: Economic Policy*, 3(3): 62-81

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. “Teacher-Student Matching and the Assessment of Teacher Effectiveness.” J. Human Resources 41 (4): 778–820.

Ehrenberg, R. G., D. D. Goldhaber, and D. J. Brewer (1995): “Do Teachers’ Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988,” Industrial and Labor Relations Review, 48(3), 547–561.

Goldhaber, Dan, and Emily Anthony. 2007. “Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching.” Rev. Econ. and Statis. 89 (1):134–50. http://www.mitpressjournals.org/doi/abs/ 10.1162/rest.89.1.134.

Hanushek, Eric A. 1971. “Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data.” A.E.R. Papers and Proc. 61 (2): 280–88.

Heckman, James J. & Kautz, Tim, 2012. "Hard evidence on soft skills," Labour Economics, Elsevier, vol. 19(4), pages 451-464.

Furnham A, Mouttafi J, Crump J. The Relationship between the Revised NEO-Personality Inventory and the Myers-Briggs Type Indicator. Social Behavior and Personality. 2003, 31(6), 577-584.

Keirsey, D., and M. Bates. 1984. Please understand me: Character and temperament types. 5th ed.

Lex Borghans & Angela Lee Duckworth & James J. Heckman & Bas ter Weel, 2008. "The Economics and Psychology of Personality Traits" Journal of Human Resources, University of Wisconsin Press, vol. 43(4).

Lundberg, Shelly, 2013. “The College Type: Personality and Educational Inequality” Journal of Labor Economics, The University of Chicago, vol. 31, no.3.

Myers, I. B. 1998. Introduction to type: Consulting Psychologists

Duckworth, A et al. “Personality Psychology and Economics,” In E. Hanushek, S. Machin, and L. Woessman, eds., Handbook of the Economics of Educa- tion , Amsterdam: Elsevier. pp. 1-181. (2011)

Myers, I., and M. H. McCaulley. 1985. Manual: A guide to the development and use of the Myers-Briggs type indicator. Palo Alto, Calif.: Consulting Psychologists Press.

Provost, J. A., and S. Anchors, eds. 1987. Applications of the Myers-Briggs type indicator in higher education. Palo Alto, Calif.: Consulting Psychologists Press.

Rimfeld, Kaili et al. "True Grit and Genetics: Predicting Academic Achievement from Personality." Journal of personality and social psychology 111.5 (2016): 780–789. PMC. Web. 7 Jan. 2018.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools and Academic Achievement." Econometrica 73 (2): 417–58.

Waude, Adam. "Conscientiousness, Personality And Behavior." *Psychologist World*, 12 Apr. 2017.

Wetzel, J. N., W. J. Potter, and D. M. O'Toole. 1982. The influence of learning styles and teaching styles on student attitudes and achievement in the introductory economics course: A case study. *Journal of Economic Education 13* (Winter): 33-39.

VIII: Appendix:

Appendix Table 1: Personality Characteristics Associated with Myer-Briggs Types

Characteristics Associated with Myers-Briggs	
How one takes in energy	
E: Extraversion	**I: Introversion**
outgoing	reflective
people person	drawn to inner world
comfortable in groups	prefer to communicate in writing
learn best by doing or discussing	learn best by reflection
broad interests	focused in depth on specific interests
drawn to take initiative in work	
How one takes in information	
S: Sensing	**N: Intuition**
solve problems by working through facts	solve problems by thinking of ideas and possibilities
prefer to use facts then form a big picture	think of the big picture first then find out the facts
trust experience over words and symbols	trust impressions, symbols, and metaphors more than experience
pragmatic	look at possibilities over realities
How one makes decisions	
T: Thinking	**F: Feeling**
enjoys thinking using logic	people and communication oriented
notices inconsistencies	concerned with harmony and nervous when it is absent
make decisions with logic and values fainess	look for what is important to others and express concern
believe truth is more important than tact	make decisions with heart and express compassion
look for logical solutions to problems	believe tact is more important than the hard truth
How one structures their outer life	
J: Judging	**P: Perceiving**
like to have things decided	prefers to stay open to whatever happens
task-oriented	spontaneous and does not like plans
like to have to-do lists	believe work is a mix of work and play

like to have work done before play	work in bursts of energy
plan work to avoid rushing at deadline	stimulated by an approaching deadline
focus intensely on goal	stay open to information so long that they may miss making important decisions

Appendix Table 2: Temperaments for Learners and Teachers

Four Temperaments for Learners and Teachers	
Dionysian: SP	
Learners:	**Teachers:**
SP learners tend to be best with hands-on learning and want to experience things for themselves. They tend to be competitive and respond well to group projects. They like to be entertained while they learned and don't do well with traditional paper work.	SP teachers represent less than 2% of higher education instructors. They are usually entertaining and rarely follow a syllabus. They value student involvement, spontaneity, and "performing" for the class and teaching through games. They are not very conscientious about grading.
Epimethean: SJ	
Learners:	**Teachers:**
SJ learners tend to do best in a lecture-based classroom. They prefer a structured learning environment and a sequential presentation of material. They do best with clear tasks and direction and tend to struggle with long independent projects and assignments that require improvising and spontaneity.	SJ teachers make up the majority of classroom teachers. They have established classroom routines and excel at the Socratic method of discussion. They do not often make exceptions for individual students. They are often found in colleges of business and education.
Promethean: NT	
Learners:	**Teachers:**
NT students tend to be independent leatners. They like to control what they study and don't need concrete examples to follow up theoretical presentations. They offten prefer to interact with the professor rather than other students and do not often enjoy group projects.	NT teachers are subject-oriented rather than student-oriented and they seek to develop the intelligence of their students. They are found mainly in higher education and tend to be inspirational to the best students but can be impersonal and impatient with those who don't show curiosity. They excel at leading problem-centered discussion and tend to move through material quickly.
Appolians: NF	
Learners:	**Teachers:**

NF students do best when they have plenty of interaction with both students and their professor. They enjoy group projects for cooperation rather than competition and learn best with case studies or discussion so that they can relate to the material. They perform best with expressive assignments such as papers rather than objective evaluation.	They have personal charisma and commitment to their students and are willing to share their own circumstances. NF teachers tend to relate to each student and are likely to adjust to the needs of students both indovidually and as a group, so they are not afraid to be behind syllabus. They tend to run democratic classrooms and bring social values into their classroom. In higher education they tend to be found in social sciences and humanities.

Appendix Table 3: Spring Grades in Calculus II

Table note: First gender listed is student, second gender listed is instructor.
Ex: Female-Male means a female students with a male instructor.

	Calculus II Final Grade			
	Female-Female	Male-Male	Female-Male	Male-Female
Fall Final Grade	0.617***	0.576***	0.586***	0.520***
	(-0.0598)	(-0.0154)	(-0.0306)	(-0.0338)
Fall Average Residual	-0.341	-0.585***	-0.461***	-0.411**
	(0.303)	(0.0753)	(0.129)	(0.173)
Extroversion-Introversion Intensity Index	0.00123	-0.0000761	-0.000475	0.000841
	(0.00258)	(0.000498)	(0.00104)	(0.00129)
Sensing-Intuition Intensity Index	0.00255	0.000884*	0.00106	-0.00142
	(0.00268)	(0.000525)	(0.00114)	(0.00139)
Thinking-Feeling Intensity Index	-0.000392	-0.0000642	-0.000103	-0.00185
	(0.00256)	(0.000569)	(0.00112)	(0.00153)
Judging-Perceiving Intensity Index	0.00106	0.00149***	0.00206*	0.00163
	(0.00254)	(0.000495)	(0.00107)	(0.00131)
EIint*Ave Fall resid	-0.00344	-0.00416*	-0.00125	0.00700
	(0.0122)	(0.00228)	(0.00442)	(0.00551)
SNint*Ave Fall resid	-0.0260**	0.000135	0.00372	0.0131**
	(0.0117)	(0.00235)	(0.00486)	(0.00609)
TFint*Ave Fall resid	-0.00713	0.00351	-0.00706	-0.00550
	(0.0111)	(0.00259)	(0.00529)	(0.00682)
JPint*Ave Fall resid	0.0229*	0.00172	0.00108	-0.000888
	(0.0118)	(0.00218)	(0.00496)	(0.00575)
_cons	0.966	0.388**	1.235***	0.871**
	(0.803)	(0.188)	(0.400)	(0.396)
N	232	4118	1002	864
R-sq	0.390	0.347	0.350	0.324

	English II Final Grade			
	Female-Female	Male-Male	Female-Male	Male-Female
Fall Final Grade	0.515***	0.419***	0.381***	0.352***

	(0.0738)	(0.0191)	(0.0375)	(0.0421)
Fall Average Residual	-0.139	-0.125**	-0.181*	-0.0160
	(0.195)	(0.0604)	(0.103)	(0.131)
Extroversion-Introversion Intensity Index	0.00195	-0.0000356	0.000902	-0.000211
	(0.00238)	(0.000486)	(0.00101)	(0.00115)
Sensing-Intuition Intensity Index	-0.00403	-0.0000812	0.0000877	-0.00103
	(0.00267)	(0.000510)	(0.00110)	(0.00128)
Thinking-Feeling Intensity Index	0.00116	-0.00104*	-0.000872	-0.00164
	(0.00246)	(0.000552)	(0.00109)	(0.00133)
Judging-Perceiving Intensity Index	0.00296	0.00180***	0.00103	0.00132
	(0.00229)	(0.000484)	(0.00104)	(0.00115)
EIint*Ave Fall resid	0.00593	-0.00133	0.00161	-0.000884
	(0.00639)	(0.00179)	(0.00347)	(0.00427)
SNint*Ave Fall resid	-0.00151	-0.00279	-0.0136***	-0.000798
	(0.00849)	(0.00189)	(0.00404)	(0.00464)
TFint*Ave Fall resid	0.00962	-0.00285	0.00535	0.000877
	(0.00800)	(0.00202)	(0.00421)	(0.00488)
JPint*Ave Fall resid	0.00527	0.00137	0.00962**	0.00149
	(0.00776)	(0.00177)	(0.00388)	(0.00438)
_cons	1.344*	0.983***	1.902***	1.685***
	(0.772)	(0.195)	(0.409)	(0.407)
N	201	3381	802	735
R-sq	0.334	0.214	0.208	0.228

Appendix Table 4: Spring Grades in English II

Table note: First gender listed is student, second gender listed is instructor.
Ex: Female-Male means a female students with a male instructor.

Appendix Table 4: Table 9 with Personality Interaction Terms

	Spring GPA					
	Calculus II	English II	Chemistry II	Physics II	History	Government
First Course final grade	**0.571*** **	**0.408*** **	**0.688*** **	**0.546*** **	**0.405*** **	**0.386*** **
	(0.0124)	(0.00874)	(0.00609)	(0.00754)	(0.0109)	(0.0106)
Ave fall Residual	**-0.469*** **	**-0.0787*** **	**-0.255*** **	**-0.368*** **	**-0.159*** **	**-0.174*** **
	(0.0580)	(0.0264)	(0.0379)	(0.0318)	(0.0349)	(0.0317)
EI Intensity Index	**0.0000627**	**0.000595*** **	**-0.000307**	**-0.000310**	**-0.0000233**	**0.000472**
	(0.000412)	(0.000227)	(0.000197)	(0.000235)	(0.000312)	(0.000300)
SN Intensity Index	**0.000826***	**-0.000422***	**-0.000156**	**-0.000458***	**0.00105*** **	**0.000185**
	(0.000439)	(0.000241)	(0.000211)	(0.000247)	(0.000330)	(0.000319)
TF Intensity Index	**-0.000477**	**-0.000652** **	**-0.000161**	**-0.000462***	**-0.000659***	**0.000170**
	(0.000465)	(0.000257)	(0.000223)	(0.000265)	(0.000353)	(0.000338)
JP Intensity Index	**0.00161*** **	**0.00186*** **	**0.00124*** **	**0.00173*** **	**0.00209*** **	**0.00127*** **
	(0.000412)	(0.000230)	(0.000200)	(0.000237)	(0.000316)	(0.000304)
EI_int*ave resid	**-0.00195**	**0.000403**	**-0.00114**	**-0.000223**	**0.000659**	**-0.00178***
	(0.00185)	(0.000825)	(0.00124)	(0.000989)	(0.00107)	(0.000987)
SN_int*ave resid	**0.00161**	**-0.00220** **	**0.00133**	**0.000126**	**-0.00100**	**-0.00176***
	(0.00194)	(0.000882)	(0.00132)	(0.00104)	(0.00113)	(0.00107)
TF_int*ave resid	**-0.000424**	**0.000433**	**-0.000372**	**0.000156**	**0.000320**	**0.00236** **
	(0.00209)	(0.000919)	(0.00139)	(0.00111)	(0.00121)	(0.00111)
JP_int*ave resid	**0.00193**	**0.00138***	**-0.00147**	**0.00234** **	**0.00149**	**0.00102**
	(0.00181)	(0.000820)	(0.00122)	(0.000993)	(0.00107)	(0.000992)
_cons	0.612***	0.421***	-0.519***	-0.235***	0.731***	0.403***
	(0.153)	(0.0790)	(0.0659)	(0.0793)	(0.103)	(0.101)
N	6249	15124	17210	15353	8669	9223
R-sq	0.339	0.266	0.572	0.429	0.286	0.289