

**CHARACTERIZING, MODELING, AND ADDRESSING
GENDER DISPARITIES
IN INTRODUCTORY COLLEGE PHYSICS**

by

LAUREN ELIZABETH KOST-SMITH

B.A., Lawrence University, 2005

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Physics

2011

This thesis entitled:
Characterizing, Modeling, and Addressing Gender Disparities in Introductory College Physics
written by Lauren Elizabeth Kost-Smith
has been approved for the Department of Physics

Noah Finkelstein

Steven Pollock

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #0607.13 and #1209.36

Kost-Smith, Lauren Elizabeth (Ph.D., Physics)

Characterizing, Modeling, and Addressing Gender Disparities in Introductory College Physics

Thesis directed by Associate Professor Noah D. Finkelstein

The underrepresentation and underperformance of females in physics has been well documented and has long concerned policy-makers, educators, and the physics community. In this thesis, we focus on gender disparities in the first- and second-semester introductory, calculus-based physics courses at the University of Colorado. Success in these courses is critical for future study and careers in physics (and other sciences). Using data gathered from roughly 10,000 undergraduate students, we identify and model gender differences in the introductory physics courses in three areas: student performance, retention, and psychological factors. We observe gender differences on several measures in the introductory physics courses: females are less likely to take a high school physics course than males and have lower standardized mathematics test scores; males outscore females on both pre- and post-course conceptual physics surveys and in-class exams; and males have more expert-like attitudes and beliefs about physics than females. These background differences of males and females account for 60% to 70% of the gender gap that we observe on a post-course survey of conceptual physics understanding. In analyzing underlying psychological factors of learning, we find that female students report lower self-confidence related to succeeding in the introductory courses (self-efficacy) and are less likely to report seeing themselves as a “physics person”. Students’ self-efficacy beliefs are significant predictors of their performance, even when measures of physics and mathematics background are controlled, and account for an additional 10% of the gender gap. Informed by results from these studies, we implemented and tested a psychological, self-affirmation intervention aimed at enhancing female students’ performance in Physics 1. Self-affirmation

reduced the gender gap in performance on both in-class exams and the post-course conceptual physics survey. Further, the benefit of the self-affirmation was strongest for females who endorsed the stereotype that men do better than women in physics. The findings of this thesis suggest that there are multiple factors that contribute to the underperformance of females in physics. Establishing this model of gender differences is a first step towards increasing females' participation and performance in physics, and can be used to guide future interventions to address the disparities.

ACKNOWLEDGMENTS

This research has been supported by the National Science Foundation (Grant No. REC 0448176, CAREER: Physics Education and Contexts of Student Learning, Grant No. 0833364, I3: Towards a Center for STEM Education, and Grant No. DRL 0910373, Understanding and Reducing the Gender Gap in Math and Science), the APS/AIP/AAPT PhysTEC program, and the University of Colorado at Boulder.

I would first like to acknowledge my thesis advisors, Noah Finkelstein and Steven Pollock. I have thoroughly enjoyed working with both of them and have appreciated their diverse views, advice, and opinions on my research. I would not have completed this thesis without their support, encouragement, and confidence in me. I would also like to thank the other members of my thesis committee, Heather Lewandowski, Paul Beale, and Diane Sieber, for their thoughtful comments and reviews. I am also grateful to the many physics faculty members and students who participated in my studies, without whom this research could not have been done.

I was fortunate to be surrounded by outstanding graduate students, post-docs, and faculty while conducting my research. My work was strengthened by the feedback and criticism that I received from the physics education research group at Colorado. Many hours of discussions with Chandra Turpen were integral in shaping my thesis work and my ideas about teaching and learning. I would also like to acknowledge colleagues in the School of Education who challenged me to think about my research in new ways, and my collaborators in psychology, Geoffrey Cohen, Tiffany Ito, and Akira Miyake.

Finally, I would like to acknowledge my incredible family. I would not have completed this thesis without the love, support, encouragement, smiles, and hugs from my husband, Zach, who has believed in me for the past five years. I would like to thank my mom and dad, Melita and Craig Kost, for their constant love and support, for teaching me to work hard to achieve my goals, and for always encouraging me to pursue my interests. My brother and sister, Adam and Caitlin, have motivated and inspired me to always be better, and have kept me smiling even from a distance. Lastly, I would like to thank my friends, Chandra Turpen, Evelyn Okimoto, and Bernadette Heyburn, for keeping me sane during graduate school. And to the Boulder softball community who I have played with, coached with, and coached for, thank you for letting me be part of your teams and giving me much needed breaks from my thesis work.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
I. Motivation	1
II. Literature Review and Large-Scale Framing	4
III. Overview of the Thesis Project	8
A. Purpose of the Thesis Project and Research Questions	8
B. Methodological Approach	10
C. Overview of the Dissertation	11
References (Chapter 1)	17
CHAPTER 2: GENDER DISPARITIES IN PHYSICS 1	20
I. Introduction	20
II. Background	21
A. Interactive Engagement	21
B. Gender Gaps in Physics	21
C. Factors That Predict Student Performance	23
III. Research Methods	24
A. Course Description	24
B. Student Population	25
C. Performance and Background Measures	26
D. Study Sample	29
IV. Results	30
A. Identifying Differences By Gender	30
B. Correlation of Student Background With Student Conceptual Performance	38

C. Estimation of the Impact of Student Background on the Gender Gap	41
V. Discussion and Conclusions	54
References (Chapter 2)	58
CHAPTER 3: GENDER DISPARITIES IN PHYSICS 2	63
I. Introduction	63
II. Background	64
III. Research Methods	66
A. Course Description	66
B. Student Population	67
C. Performance and Background Measures	68
D. Study Sample	70
IV. Results	71
A. Tracking Students From Physics 1 to Physics 2	71
B. Identifying Differences By Gender	77
C. Correlation of Student Background With Student Conceptual Performance	88
D. Estimation of the Impact of Student Background on the Gender Gap	90
V. Discussion and Conclusions	99
References (Chapter 3)	104
CHAPTER 4: SELF-EFFICACY IN PHYSICS	108
I. Introduction	108
II. Background	109
A. Self-Efficacy	109
B. Sources of Self-Efficacy	112

C. Gender Differences in Self-Efficacy	114
D. Self-Efficacy in Physics	116
III. Research Methods	118
A. Physics Self-Efficacy and Identity Survey	118
B. Survey Sample	123
C. Statistical Methodology	127
IV. Results	129
A. Students' Sense of Physics Self-Efficacy	129
B. Sources of Self-Efficacy	145
C. Regression Modeling	153
V. Conclusions	164
VI. Discussion	165
References (Chapter 4)	169
CHAPTER 5: IDENTITY IN PHYSICS	174
I. Introduction	174
II. Background	175
A. Learning as Identity Formation	175
B. Model of Physics Identity	178
III. Research Methods	180
IV. Results	182
A. Students' Sense of Self-Identity	183
B. Students' Sense of Recognition by Others	194
C. Regression Modeling	201

V. Discussion and Conclusions	204
References (Chapter 5)	208
CHAPTER 6: IDENTITY THREAT AND A SELF-AFFIRMATION INTERVENTION IN PHYSICS 1	210
I. Introduction	210
II. Background	211
A. Identity Threat	212
B. Self-Affirmation	214
III. Research Methods	215
A. Study Design	215
B. Course Description	219
C. Study Sample	219
D. Values Affirmation and Control Writing Exercises	221
E. Stereotype Endorsement Measure	222
F. Outcome Measures	224
G. Statistical Methodology	225
IV. Results	227
A. Exams	228
B. Course Grade	232
C. Force and Motion Conceptual Evaluation	234
D. Historical Comparison	237
E. Other Course Components	239
F. Replication Study	240
V. Discussion and Conclusions	245

References (Chapter 6)	250
CHAPTER 7: CONCLUSIONS	255
I. Summary of Findings	255
II. Implications	263
III. Directions for Future Research	265
References (Chapter 7)	269
BIBLIOGRAPHY	271
CHAPTER 2 APPENDIX	288
CHAPTER 3 APPENDIX	290
CHAPTER 4 APPENDICES	292
A. Physics Self-Efficacy and Identity Survey	292
B. Factor Analysis Results	299
C. Summary Tables of Survey Statistics	302
D. Regression Models for Physics 2	306
CHAPTER 5 APPENDICES	308
A. Summary Tables of Survey Statistics	308
B. Regression Models for Exams in Physics 1 and Physics 2	311
CHAPTER 6 APPENDICES	312
A. Affirmation and Control Exercises and TA Script	312
B. Attitudes Towards Science Survey	322
C. Table of Raw and Adjusted Average Values of Outcome Measures	327
D. Regression Models for Each Exam	329
E. Regression Models for Homework, Participation, and Clicker Scores	331

CHAPTER 1: INTRODUCTION

I. Motivation

In response to growing concern over the United States' future economic prosperity and national security, the National Academies set out to investigate where the US fit into the global science and technology enterprise, and to make recommendations to improve their position. In their 2007 report, *Rising Above the Gathering Storm* [1], the National Academies argued that in order to compete globally the US must be a leader in science and technology, which requires a knowledgeable and scientifically literate workforce. To that end, they put forth four recommendations, two of which involve increasing the pool of people prepared for careers in science and technology. One recommendation was to improve mathematics and science education at the K-12 level in order to increase the number of students who are prepared to study science, technology, engineering, and mathematics (STEM) in college; while the other was to develop and retain the best and brightest students and scientists by increasing the number of US students earning undergraduate and graduate degrees in STEM fields. The National Academies assert that increasing the number and quality of students trained for a career in STEM fields is critical for meeting the workforce needs of the US in the 21st century and protecting the prosperity, health, and security of the nation.

While the number of students earning bachelor's degrees in science and engineering fields is increasing, the percentage has declined from about 36% in the late 1960's to 32% in 2006 [2]. Master's degrees have declined similarly (from 29% to 21%), though doctoral degrees earned in science and engineering remains high (around 65%). In physics, the number of

bachelor's degrees awarded lags behind most other science and engineering fields. In 2008, 2% of the bachelor's degrees awarded in science and engineering were in physics, lower than biology (33%), engineering (28%), mathematics (6%), and chemistry (5%) [3,4]. In response to the National Academies report [1] and the American Competitiveness Initiative [5] (which later became the America COMPETES Act [6]), the American Physical Society and the American Association of Physics Teachers launched a joint initiative in 2007 to double the number of bachelor's degrees earned in physics. This doubling initiative was proposed, "to address critical national needs including K-12 education, economic competitiveness, energy, security, and an informed electorate," [7]. In their mission statement, they cite areas of need and of documented shortages of physics-educated professionals, such as biophysics, nanoscience, medical physics, nuclear physics, and secondary physics teachers [8]. To meet the need for a scientifically literate US workforce, we need to increase the number of students earning degrees in physics, and female students represent a potential untapped resource to fill this need.

According to the American Institute of Physics [9], 46% of students who took high school physics in 2001 were female, up from 39% in 1987. Though encouraging that nearly half of high school physics students are now female, there is some evidence that males and females are taking different high school physics courses. Of students who took the calculus-based AP (Advanced Placement) physics test in 2004, 25% were female, and 35% of students who took the algebra-based AP physics test were female [10]. These numbers suggest that females are less likely to take AP physics courses, and when they do, are more likely to take the algebra-based over the calculus-based course. Fewer females are choosing to take the rigorous high school physics courses that are recommended to prepare them for the study of physics at the college level.

Turning to students at the undergraduate and graduate level, in 2008, females earned 38% of the bachelor's degrees and 33% of the doctoral degrees awarded in the sciences [3]. Looking at individual disciplines, females represent 60% of the bachelor's degrees awarded in biology, 50% in chemistry and 44% in mathematics. Despite reaching parity of males and females in these disciplines, females still earn only 20% of the bachelor's degrees, 21% of the master's degrees, and 19% of the doctoral degrees awarded in physics. Physics has one of the lowest representations of females, and is comparable to representation in computer science and engineering (both 19% female at the undergraduate level). These national trends are reflected at the University of Colorado at Boulder (CU). Females make up only 25% of the students who enroll in introductory physics and about 15% of physics majors. This under-representation of females in physics is cause for concern [1,11].

One might ask: Why do we need to increase the participation of females in physics? There are several arguments, one of which was given above. There is a need for more and better-trained scientists to secure the nation's future prosperity, and female students are an, as yet, under-utilized resource [12,13]. Females make up half of the talent pool from which to draw students to physics, and yet they make up only 20% of those students who choose to pursue careers in physics. Further, diverse groups are more productive and successful than uniform groups, and females can potentially bring new perspectives to physics [14,15,16]. Lastly, degrees in science and physics lead to more lucrative and prestigious career opportunities. Females should have the same opportunities to earn a degree in physics and pursue those careers as males, but research shows that they currently do not [12]. Females, more than males, face discrimination, implicit biases in evaluative situations, and organizational structures that put them at a disadvantage, at all steps along the academic and career ladder [12].

To increase the numbers of females participating in physics, we first need to understand the factors that influence whether or not females choose to pursue a career in physics. A critical first step towards physics (and other science) careers is the introductory physics course, which is traditionally seen as a “gate-keeping” course. Success in this course is a requirement for pursuing a physics (or science) degree, and as such, we focus our studies on the calculus-based Physics 1 and Physics 2 courses, which are for physics majors and engineers. Understanding the factors that do and do not promote student learning in these courses is crucial to understanding the gender gap and finding ways to eliminate it. In this thesis, we characterize gender differences in student performance, retention, and psychological factors in the introductory physics courses, as well as what factors are correlated with the differential performance of males and females. Identifying these factors is a first step in clarifying the mechanisms by which gender differences are established and guides the development of interventions to address the disparities. Based on these findings, we implemented a psychological intervention, and tested its effectiveness in reducing gender differences in performance.

II. Literature Review and Large-Scale Framing

In this section we broadly summarize the research on gender issues in science education in order to frame and motivate the work we present in this thesis. Literature that is more specifically relevant to individual studies will be discussed in the following chapters.

Much of the work on gender issues in science education has been to either identify individual characteristics of males and females that lead to their differential participation and performance in science, or to research, develop, and implement interventions to increase females’ interest or performance in science. In a sense, research efforts either focus on the individual or on the practices that take place in science classrooms. This division of the research

is clear in Kahle and Meece's [17] summary of the research on gender issues in science, which has two major sections, one on the factors that contribute to differential achievement and participation of males and females and one on intervention programs to address the differential achievement and participation. This organization of the research, as studying individual characteristics or classroom practices, is also presented in other summaries of the literature on gender issues in science [18,19,20].

Research on individual characteristics identifies several variables that impact performance and retention of males and females in science. Many researchers have investigated differences in the cognitive abilities of males and females, including spatial reasoning ability and mathematics reasoning ability. Results from meta-analyses suggest that gender differences in spatial reasoning ability are small, and do not explain the differences in performance or participation in science [21,22,23]. Differences in mathematics ability are not consistent. While males consistently outscore females on standardized mathematics tests, when course grades are used to measure mathematics achievement, the gender differences are nonexistent or favor females [17,19]. In addition to individual cognitive variables, attitudinal variables have also been investigated. These include students' interest and enjoyment in doing science, anxiety related to science, and students' perceptions of their science ability. Two recent meta-analyses on student attitudes towards science find that females have less positive attitudes towards science than males [24,25]. Other individual characteristics of males and females that have been investigated are home and family variables, including the level of parents' education, socioeconomic status, parental attitudes towards science, parental encouragement, opportunities for engaging in science activities at home, and gendered expectations of parents. Kahle and Meece [17] summarize the findings by reporting, "in general, the socialization experiences girls receive in the home are not

likely to encourage achievement in science,” [p. 549]. Researchers have also suggested that males and females have different ways of knowing [26], in particular, females often try to make connections between themselves and what they are learning, an approach that Belenky *et al.* argue is not valued in traditional education. Through all of this work, there is a focus on the individual characteristics of males and females, how they differ, and how they contribute to the differential performance and participation in science.

The other major focus of research on gender issues in science has been on the classroom environment and classroom practices, and how these impact the participation and performance of males and females in science. Researchers have investigated teacher expectations and behavior, how different types of instruction impact males and females, and the structure of student interactions [17]. In one study, teachers reported having equal expectations for male and female students, but researchers observed teachers treating students in ways that suggested they did not have the same expectations for males and females [27]. Many studies have shown that female students prefer cooperative, over competitive, learning environments and may learn better from inquiry-based activities [17,20]. Further, several studies have documented gender differences in how students engage when working with groups and when working with laboratory equipment [28,29,30]. Work has also examined the science interests of males and females in order to adapt the curriculum so that it is more in line with the interests of females [20]. This leads to research on interventions, or on how classroom practices and environments can be modified in ways that enhance the performance and participation of females. Kahle and Meece [17] summarize the goals of most interventions as, “1) demasculinizing and demystifying science, usually by exposure to role models and career information; 2) improving girls’ self-confidence and self-perceptions of their ability to do science; 3) implementing teaching strategies that actively

involved girls in science lessons; and 4) developing girls' skills of doing science," [p. 550]. They conclude that while interventions have been successful at identifying what factors impact females' confidence and retention in science, they have been less successful at identifying what factors impact the achievement gap between males and females.

As described above, research on gender issues in science can roughly be categorized as research related to individual factors and research related to classroom practices (including interventions, which modify classroom practices). What is less represented in the literature is work that bridges characteristics of individuals with the practices that they engage in. Said another way, little work has been done to investigate the individual cognitive and psychological factors that influence student performance and participation in particular environments as they engage in specific practices, and to develop interventions based on those findings to implement in the environments that have been studied. In their 2007 summary, Scantlebury and Baker [19] concluded that, "intervention programs designed to increase girls' participation in science had limited impact, because they focused on single rather than multiple causes and were not grounded in theoretical models that integrated psychological and sociocultural variables," [p. 263]. It is clear from the prior literature on gender issues in science that there is no single factor that can explain the differences in participation and performance between males and females in science. To understand gender differences in participation and performance in physics, our goal in this thesis, multiple factors need to be considered, and their impact tested in specific learning environments. Once the factors that impact performance in that particular learning environment are well understood, interventions to address the gender differences can be developed and explored.

III. Overview of the Thesis Project

A. *Purpose of the Thesis Project and Research Questions*

The goal of our work is to understand the underrepresentation and underperformance of females in physics compared to males, and to find ways to address the disparities. As described above, we focus on the introductory physics courses, as success in these courses is required for further study in physics. Because students who enroll in these courses are primarily science or engineering majors, they likely already have an interest in science and likely have the preparation and background needed to be successful in science or engineering majors. The students in these courses make up the pool from which potential physics majors can be drawn. It is possible that intervening in these early courses will help retain more females who are already interested in pursuing physics, as well as recruit females into physics.

Responding to calls in the literature for interventions that are theoretically grounded, in this thesis we seek to bridge individuals and practices. We study characteristics of individuals in specific environments as they engage in particular practices. We then use what we learn about individual differences to develop and implement an intervention to address the gender differences in student performance. We begin by identifying and modeling gender differences in the introductory physics courses. We focus on gender differences in three areas: student performance in the course (both in terms of course grades and conceptual survey performance), retention from Physics 1 to Physics 2, and students' attitudes and beliefs about physics and about learning physics. We not only characterize these differences, but also model what academic background factors of students contribute to the differences we observe in Physics 1 and Physics 2. We then shift the focus to the underpinning psychological factors that contribute to student learning and performance in physics. We explore not only gender differences in physics self-

confidence and students' identification with physics, but also how these psychological factors correlate with student performance in the introductory courses. In this way, we work towards building a model of student performance in the introductory physics courses that draws on a variety of factors, including cognitive, social, and psychological factors. Finally, informed by the results of these prior studies, we implement a psychological intervention aimed at enhancing female students' performance in the introductory physics course.

In this dissertation, we address the following research questions:

1. On what measures, and to what extent, do we observe differences by gender in the introductory physics courses? For example, are there gender differences in conceptual learning, course grades and their components, attitudes and beliefs, and prior knowledge and preparation?
2. Can gender differences in student performance be accounted for by factors other than gender, namely, background differences of males and females?
3. What role do underlying psychological factors, such as students' physics self-confidence (self-efficacy) and sense of being a physics person (physics identity), play in student performance and in the gender differences in student performance?
4. Does identity threat (the fear of being devalued based on a group identity) negatively impact females' performance in the introductory physics course? Can the performance differences between males and females be reduced or even eliminated with a psychological, self-affirmation intervention?

B. *Methodological Approach*

The methodological approach employed in this thesis is primarily quantitative (though some qualitative methods are used to analyze students' open-responses to survey questions). In Chapters 2 through 5 an epidemiological approach is used in which data are collected and analyzed from roughly 10,000 undergraduate students enrolled in the introductory physics courses over a seven-year time span at CU. These analyses are used to identify key factors found to strongly correlate with student performance and gender differences in student performance. Appropriate linear modeling techniques are applied. A variety of student factors are considered including academic background factors (e.g. standardized test scores, high school courses taken, high school GPA, and conceptual pre-test scores), prior attitudes and beliefs about physics, physics-related self-confidence, and identification with physics. Data on each of these factors are collected either from university records or self-reported by students on pre- and post-course surveys. Data on student performance in the course were provided by the instructors in each course. We seek to model the effects of a broad range of cognitive, social, and psychological variables on students' performance in the introductory physics courses at CU.

In Chapter 6, a randomized experimental design is used to test the impact of a self-affirmation intervention on student performance. Students from the same course are randomly assigned to complete either a treatment or control exercise, thereby ensuring that differences in the performance between the two groups could reasonably be attributed to the intervention. More details about the methodological approach used in this, and all, studies will be provided in the following chapters.

Researchers have used a variety of methodologies to investigate the underrepresentation of females in physics and the sciences, everything from large-scale, national surveys of students

to deep ethnography that focuses on understanding the experiences of a few individuals. In this thesis, we try to strike a balance between these two methodologies. We attempt to capture the experiences of individual students (through surveys), but also take advantage of the large numbers of students that take the introductory physics courses (using statistical analysis techniques). While we lose information about the choices and experiences of individual students, we gain the ability to identify trends across populations of students, and therefore to predict what kinds of interventions will be beneficial to the majority of the students. To be explicit, we did not: conduct interviews with students, track individual students as they make their way (or don't) through the entire physics major, observe the physics courses to identify practices of individual instructors that may or may not encourage females to pursue physics, study the culture of practicing physicists to understand how it is friendly or hostile towards females, or observe individuals or groups of students as they engage (or don't) in the physics course. While we believe these are all worthwhile directions to pursue in understanding why females are under-represented in physics, they are not taken up in this thesis. Further, the work in this thesis will not provide an answer for what should be done about the lack of female representation in physics. Based on our findings, we can propose changes to the introductory physics courses that may benefit female students, but these suggested changes will require further research. We feel that this work is an important step towards understanding gender disparities in physics and eventually finding ways to address them.

C. Overview of the Dissertation

Here, I give an overview of this thesis, briefly describing the studies and findings of each chapter. As stated above, the relevant prior literature and methodological approaches used for each of the studies in this thesis will be discussed in detail in the following chapters.

In Chapter 2, we characterize gender differences in the first-semester, calculus-based mechanics course (Physics 1), and determine which factors are correlated with differential performance at our institution. We first identify the measures on which we observe gender differences, including both measures of student learning in the course and background factors. We then determine to what extent differences in student background contribute to gender differences in performance that we observe at the end of the course. We identify gender differences in several aspects of the introductory course. Significant gender differences exist on both the pre- and post-course conceptual survey of physics. Additionally, males outscore females on in-class exams. However, because females outscore males on homework and participation, the total course grades of males and females are not significantly different in any individual semester (though averaging over all semesters reveals a small, but significant, gender difference favoring males). Finally, females have significantly less favorable attitudes and beliefs towards physics than males at both the beginning and end of the Physics 1 course. When these differences in prior understanding and attitudes are controlled for, meaning we only compare students with the same background, the gender gap is reduced from about 11% to about 3%. The results indicate that the gender gap exists, even in interactive physics classes, at our institution, but is largely associated with differences in previous physics and math knowledge and incoming attitudes and beliefs.

In Chapter 3, we continue to characterize gender differences in participation and performance by examining the second-semester, calculus-based electricity and magnetism (E&M) course (Physics 2). This course is different from Physics 1 in many ways; in particular, fewer students (male or female) have significant exposure to the E&M content before coming to the course. By studying gender differences in the E&M course, we are able to tease apart some

of the similarities and differences between performance in Physics 1 and Physics 2, and get a sense of which, if any, contextual factors differentially impact males and females. We find no differences in the retention rates of males and females from Physics 1 to Physics 2 for students overall; males and females are equally likely to continue or not continue from Physics 1 to Physics 2. However, there are small differences in the retention rates of male and female physics majors, with males systematically more likely to continue and less likely to drop out than females. The trends in male and female course grades in Physics 2 match those observed in Physics 1, except for some notable cases in which males and females have significantly different course grades. Despite males and females having similar E&M conceptual pre-test scores at the beginning of Physics 2, males outperform females at the end of the semester by about 6%. This post-test gender gap can largely be attributed to differences in males' and females' prior physics performance, mathematics standardized test performance, and students' attitudes and beliefs. A multiple regression model of students' conceptual performance suggests these prior factors can account for up to 60% of the observed gender differences. Taken all together, the study of the second-semester physics course indicates that there is not one single factor that can explain the under-representation of females in physics, but it is rather the building up of small differences between males and females over time that may be responsible for the large disparities in participation of males and females in physics.

In Chapter 4, we turn to the psychological factors that underpin student performance in Physics 1 and Physics 2. We first explore students' sense of physics self-efficacy, or their confidence related to completing the tasks necessary for success in physics, and the sources of self-efficacy, those experiences that contribute to the development of students' self-efficacy. Prior literature suggests that self-efficacy is an important factor in student performance and

retention in science. We find that males have significantly higher physics self-efficacy than females. Additionally, these self-efficacy beliefs are useful predictors of students' conceptual survey and exam performance in Physics 1 and Physics 2. For students in this study, background measures of physics and mathematics understanding account for about 70% of the gender gap in conceptual survey post-test scores, and students' self-efficacy beliefs account for an additional 12% of this gender gap. Also, there are significant differences in student responses to questions about the four sources of self-efficacy, with males giving more positive interpretations of their experiences than females for each source. Each of the four sources of self-efficacy is a significant predictor of students' post-course self-efficacy in addition to students' pre-course self-efficacy and their performance in the course. Results from the study of students' physics self-efficacy suggest that it is an important factor that contributes to student performance in the course, and providing opportunities for males and females to positively develop their self-efficacy may be important for reducing gender disparities in the physics courses.

In Chapter 5, we explore another underpinning psychological factor, students' sense of physics identity, both how they recognize themselves and how they perceive being recognized by others as a physics person. Research on gender issues in science points towards identity as a critical factor to consider when trying to understand why females choose to participate or not in science. We find that males are significantly more likely to self identify and to report that they are recognized by others as a physics person. When asked why they do or do not see themselves as a physics person, most student responses involve either their interest in physics or their abilities in physics (both are aspects of physics identity as defined by prior work). More than half of both males and females report that they felt like less of a physics person at the end of Physics 1 or Physics 2 compared to the beginning of the class. Again, the most common reasons were

loss of interest in physics or because they didn't do well in physics. Regression modeling was used to determine whether students' physics identity was a useful predictor of student performance in the course. Though self-identity was a useful predictor on its own, when measures of students' self-efficacy were included, self-identity was no longer a significant predictor of end-of-semester conceptual survey performance. Though identity was not a useful predictor of student performance in the course, we suspect that it is an important factor that impacts whether or not students choose to pursue physics. Students' decrease in their physics identity over the course of Physics 1 and Physics 2 is disconcerting and needs to be further researched.

In Chapter 6, we report the results of a large-scale, randomized classroom study to test the effectiveness of a self-affirmation intervention in reducing gender difference in performance in Physics 1. In the course of this research, we hypothesized that identity threat may be inhibiting females' performance in the introductory courses. Researchers have demonstrated that identity threat can be alleviated through self-affirmation. We find that the self-affirmation intervention was successful in reducing the gender gap both in students' in-class exam scores and their conceptual post-test scores. The reduction in the gender gap was primarily due to affirmed females' increased performance compared to females in a control group. Further, the effect of the intervention was moderated by students' endorsement of the stereotype that men are expected to do better in physics than women. There was a negative relationship between stereotype endorsement and performance on exams and the conceptual survey for females, unless they had completed the values affirmation exercise, in which case they were buffered from the negative impact of stereotype endorsement. In addition to the results from the initial study, we also report the results from a replication study. We did not find a statistically significant effect of the

intervention on the gender gap in the replication study. We discuss implications of both the original and replication study results for future implementations of the self-affirmation intervention.

In Chapter 7, we summarize the findings across all chapters, discuss the implications of these results, and recommend directions for future research. Taken all together, we find that there are many factors that impact the gender differences in performance and retention in physics. The model that we have built of student performance in the introductory physics courses includes background cognitive factors, such as prior physics and mathematics understanding, prior psychological factors, such as attitudes and beliefs about physics and students' physics self-efficacy, and social factors, such as gender and identity threat. In modifying classroom environments and practices to increase the participation and performance of female students, all of these factors need to be addressed.

References (Chapter 1)

- [1] National Academy of Sciences, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Future* (National Academy Press, Washington, DC, 2005).
- [2] National Science Foundation, Division of Science Resources Statistics, *S&E Degrees: 1966 – 2006*, NSF 08-321 (Arlington, VA, 2006). Available from <http://www.nsf.gov/statistics/degrees>.
- [3] National Science Foundation, Division of Science Resources Statistics, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2011*, NSF 11-309 (Arlington, VA, 2011). Available from <http://www.nsf.gov/statistics/women/>.
- [4] In NSF's reported statistics, they include psychology and social sciences as part of science and engineering. In our reported percentages, we do not include psychology and social sciences as part of science and engineering.
- [5] Domestic Policy Council, Office of Science and Technology Policy, *American Competitiveness Initiative: Leading the World in Innovation* (2006). Available at www.nsf.gov/attachments/108276/public/ACI.pdf.
- [6] America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science (COMPETES) Act, Public Law 110-69, 121 Stat. 572 (2007). And, America COMPETES Reauthorization Act of 2010, Public Law 111-358, 124 Stat. 3982 (2011).
- [7] APS/AAPT Doubling Initiative Mission Statement (2007). Available at <http://www.aps.org/programs/education/undergrad/faculty/double-mission.cfm>.
- [8] See for instance: B. Thomadsen, The shortage of radiotherapy physicists, *Journal of the American College of Radiology* **1**(4), 280 (2004); APS Panel on Public Affairs, Committee on Energy and Environment, *Readiness of the U.S. Nuclear Workforce for 21st Century Challenges* (2008). Available at <http://www.aps.org/policy/reports/popa-reports/index.cfm>; S. White and C. L. Tesfaye, *Who teaches high school physics? Results from the 2008-09 nationwide survey of high school physics teachers* (AIP Statistical Research Center, College Park, MD, 2010). Available from www.aip.org/statistics.
- [9] R. Ivie and K. N. Ray, *Women in Physics and Astronomy, 2005*, AIP Report No. R-430.02 (College Park, MD, 2005). Available from www.aip.org/statistics.
- [10] College Board, *Advanced Placement Summary Information* (The College Board, New York, 2005).

- [11] National Science Board, *The Science and Engineering Workforce: Realizing America's Potential*, NSF 03-69 (National Science Foundation, Arlington, VA, 2003).
- [12] National Academy of Science, *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* (National Academy Press, Washington, DC, 2006).
- [13] American Physical Society, Committee on the Status of Women in Physics, *Gender Equity Report* (unpublished, 2007). Available from <http://www.aps.org/programs/women/workshops/gender-equity/>.
- [14] L. Schiebinger, *Has Feminism Changed Science?* (Harvard University Press, Cambridge, MA, 1999).
- [15] N. Brickhouse, Bringing in the outsiders: Reshaping the science of the future, *Journal of Curriculum Studies* **26**(4), 401 (1994).
- [16] J. Kenway and A. Gough, Gender and science education in schools: A review "with attitude," *Studies in Science Education* **31**(1), 1 (1998).
- [17] J. B. Kahle and J. Meece, "Research on gender issues in the classroom" in *Handbook of Research on Science Teaching and Learning*, edited by D. Gabel (National Science Teachers Association, Washington, DC, 1994) p. 542.
- [18] J. Oakes, "Opportunities, achievement, and choice: Women and minority students in science and mathematics" in *Review of Research in Education*, edited by C. Cazden (The American Educational Research Association, Washington, DC, 1990) p. 153.
- [19] K. Scantlebury and D. Baker, "Gender issues in science education research: Remembering where the difference lies" in *Handbook of Research on Science Education*, edited by S. Abell and N. Lederman (Lawrence Erlbaum, Mahwah, NJ, 2007) p. 257.
- [20] J. S. Brotman and F. M. Moore, Girls and science: A review of four themes in the science education literature, *Journal of Research in Science Teaching* **45**(9), 971 (2008).
- [21] M. C. Linn and A. C. Peterson, "Facts and assumptions about the nature of sex differences" in *Handbook for Achieving Sex Equity Through Education*, edited by S. S. Klein (Johns Hopkins University Press, Baltimore, MD, 1985) p. 53.
- [22] M. C. Linn and J. S. Hyde, Gender, mathematics, and science, *Educational Researcher* **18**(8), 17 (1989).
- [23] J. S. Hyde, S. M. Lindberg, M. C. Linn, A. B. Ellis and C. C. Williams, Gender similarities characterize math performance, *Science* **321**, 494 (2008).

- [24] M. Weinburgh, Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991, *Journal of Research in Science Teaching* **32**, 387 (1995).
- [25] J. Osborne, S. Simon and S. Collins, Attitudes towards science: A review of the literature and its implications, *International Journal of Science Education* **25**, 1049 (2003).
- [26] M. F. Belenky, B. M. Clinchy, N. R. Goldberger and J. M. Tarule, *Women's Ways of Knowing: The Development of Self, Voice, and Mind* (Basic Books, New York, 1986).
- [27] J. B. Kahle, A. Anderson and A. Damjanovic, A comparison of elementary teacher attitudes and skills in teaching science in Australia and the United States, *Research in Science Education* **21**, 208 (1991).
- [28] J. Jovanovic and S. King, Boys and girls in the performance-based science classroom: Who's doing the performing? *American Educational Research Journal* **35**, 477 (1998).
- [29] A. Kelly, The construction of masculine science, *British Journal of Sociology of Education* **6**, 133 (1985).
- [30] M. G. Jones, L. Brader-Araje, L. Carboni, G. Carter, M. Rua, E. Banilower, et al., Tool time: Gender and students' use of tools, controls, and authority, *Journal of Research in Science Teaching* **37**, 760 (2000).

CHAPTER 2: GENDER DISPARITIES IN PHYSICS 1

I. Introduction

In this chapter, we seek to characterize the gender gap in introductory physics, and those factors that are correlated with the differential performance at our institution. Identifying these factors is a first step in clarifying the mechanisms by which the gender gap is established and will lead to future work that guides interventions to address the disparity. We address the following research questions: 1) On what measures do we observe differences by gender in the introductory physics course, for example, conceptual learning, components of the course grade, attitudes and beliefs, and prior knowledge and preparation? 2) Are measures of background correlated with student performance in the course (as measured by a conceptual learning survey) and correlated with gender? 3) To what extent do differences in males' and females' backgrounds contribute to the post-course gender gap in Physics 1?

We first summarize the prior findings on the benefits of interactive engagement for student conceptual understanding, gender gaps in physics, and factors that influence student performance in physics. We then outline the research methods that were used for this study including a description of the Physics 1 course, the student population in this course, the measures of performance and student background that were examined, and the sample of students used in this study. We then present the results of our study in three sections, corresponding to the three research questions set out above. We find that there are several aspects of the introductory course in which we identify gender differences, including males' and females' performance on a conceptual survey and course grades, prior physics and mathematics

understanding, and their attitudes and beliefs about physics. When these differences in prior understanding and attitudes are controlled for, meaning we only compare students with the same background, the gender gap is reduced from about 11% to about 3%. The results indicate that the gender gap exists in interactive physics classes at our institution, but is largely associated with differences in previous physics and math knowledge and incoming attitudes and beliefs.

II. Background

A. Interactive Engagement

Prior research has consistently demonstrated the benefits of using interactive engagement (IE) techniques during instruction [1,2,3]. Hake [1] defined interactive engagement methods as those, “designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors,” [p. 65]. His survey of traditional and interactive courses found that interactive engagement courses had average normalized learning gains,

$$\langle g \rangle = \frac{\langle post \rangle - \langle pre \rangle}{100 - \langle pre \rangle},$$

almost two standard deviations higher than average gains in traditional courses. At the University of Colorado, classes that use IE techniques [4,5] have average normalized learning gains on conceptual assessments that range from 32% to 64%.

B. Gender Gaps in Physics

While the use of interactive engagement techniques has been shown to facilitate learning for both males and females, some research has suggested that females may benefit more from an interactive pedagogy than males [6,7]. Additionally, many of the recommendations made for increasing the participation of females in physics and in the sciences [8,9,10,11,12] align with IE

techniques. This suggests that using interactive engagement techniques may help reduce, or even eliminate, gender differences in performance that have been observed in introductory physics courses [13,14]. Indeed, this is what researchers at Harvard University found [14]. They examined the pre- and post-course gender gaps on the Force Concept Inventory (FCI) [15] in seven offerings of their introductory, calculus-based physics course for non-majors. This course is about one-third female. The seven semesters in the study varied in their level of interactive engagement. For the first semester, the course was taught using traditional, lecture methods. In the next four semesters, the lectures incorporated Peer Instruction and ConcepTests [4], but the recitation sections remained traditional with teaching assistants lecturing to students. In the final two semesters of the study, Peer Instruction was used during the lectures, and recitations were reformed to include small group work around the *Tutorials in Introductory Physics* [5] and problem-solving sessions where students worked on Context Rich Problems [16,17]. Lorenzo, *et al.* found that as the course became more interactive, the post-course gender gap decreased, to the point that for the semesters in which all of the interactive engagement techniques were used, the post-course gender gap was not statistically significant. This was the case despite a significant pre-course gender difference in all semesters. In the fully interactive semesters, both males and females had significant gains, but females had slightly larger gains, resulting in males and females having about equal post-test scores.

Despite these encouraging results, elsewhere [13,18,19,20], the gender gap has persisted despite the use of interactive engagement techniques. Docktor and Heller [19] report that over forty offerings of their introductory, calculus-based physics course for scientists and engineers, which uses Cooperative Problem Solving and Context Rich Problems [16,17], the average FCI pre-test gender gap was about $15.3 \pm 0.5\%$ and the post-test gender gap was $13.4 \pm 0.6\%$. Brewe *et*

al. [20] compared post-course FCI gender differences for lecture courses versus courses that used Modeling Instruction [21], a curriculum in which students are focused on building, validating, and deploying scientific models. The pedagogy of the course includes several interactive features, such as an emphasis on conceptual understanding, students working cooperatively in groups, and the use of multiple representational tools. While both males and females had significantly higher post-course FCI scores in the Modeling Instruction courses compared to the lecture courses, the gender gap in post-FCI scores was higher for students in the Modeling course than for students in the lecture course (20.2% versus 13.5%). While it is clear that IE methods improve student learning gains, it is less apparent that IE techniques eliminate the gender gap.

C. Factors That Predict Student Performance

The fact that there is often a gender gap in the pre-test scores (before any instruction), suggests that there are differences in preparation between male and female students when they enter the introductory physics course, and we suspect these differences may contribute to the persistence of the gender gap. Several researchers have investigated the background factors that influence student performance in introductory physics. Hazari *et al.* [22] found that mathematics preparation was a significant predictor of students' college physics grade. Others have found an influence of high school physics experience on college physics performance. Sadler and Tai [23] find that taking a high school physics course is positively related to college physics course grade, even when controlling for students' self-reported academic and demographic background. Furthermore, the pedagogy of the high school class is related to a student's college performance. Students who take high school classes that cover fewer topics in more depth have higher grades in a calculus-based college course than students whose high school classes cover more topics in less depth (the difference is almost a full letter grade) [24]. Affective factors, such as father's

encouragement and family beliefs about science, have also been shown to influence student performance [22]. These studies have all focused on students' course grades as the measure of student performance. Less work has been done on the factors that influence student performance on research-based conceptual learning surveys. As stated above, the curriculum and level of engagement in the course influences student conceptual learning [13,14,20]. Meltzer [25] found that while students' conceptual pre-test scores were not correlated with their normalized learning gains on the Conceptual Survey of Electricity (CSE), students' pre-instruction mathematics skill was correlated with CSE gains, suggesting that differences in learning gains between two populations may be due in part to different incoming math skill, rather than different pre-test scores or abilities to learn physics concepts.

III. Research Methods

A. Course Description

The data in the following studies were collected from seven offerings (Spring 2004 to Spring 2007) of the first semester, calculus-based introductory mechanics course (Physics 1) at the University of Colorado. These are large-enrollment courses that typically have 400 to 600 students. Each semester was taught by a different instructor, and all seven instructors were male [26]. Each semester included in the study will be identified by a code. All codes will have a form similar to A1-2, where the letter represents the lead instructor, the 1 indicates that this is a semester of Physics 1, and the 2 indicates that it is the second time that Professor A has taught this course in the semesters included in our data set. All seven classes used interactive engagement (IE) techniques, some to a higher degree than others. Each of the seven classes employed student discussions around conceptual questions (Peer Instruction [4]) in lecture, online homework systems [27], and voluntary help-room sessions on problem-solving

homework. Four of the seven classes used *Tutorials in Introductory Physics* [5] and Learning Assistants [28] during a one-hour per week recitation, while the remaining three classes held more traditional recitation sections. There is no laboratory associated with this course. A more detailed description of the course structure can be found in previous work [29]. We categorize the three classes that held traditional recitation sections (A1-1, B1-1, and C1-1) as IE 1 (partially interactive) and the four classes that used *Tutorials* during recitation sections (D1-1, E1-1, F1-1, and G1-1) as IE 2 (fully interactive). Our definitions of IE 1 and IE 2 classes are similar, but not identical, to the definitions used in prior studies [14]. For instance, implementations of IE 2 curricula at Harvard University included cooperative problem solving activities that were not part of IE 2 courses at CU. Additionally, while we categorize classes as either IE 1 or IE 2, we also recognize that there are a variety of faculty teaching these classes who have differing levels of experience and familiarity with the interactive engagement methods that are employed. Though the curriculum looks the same, we know that how the curriculum is enacted can be very different [30].

B. *Student Population*

The student population in the Physics 1 course is about one-quarter female. About half of the students are engineering majors and about 20% are other science majors. Only about 6% of the students who enroll in introductory physics are declared physics majors. There are some differences in the distributions of student major for males and females, but the same percentage of males and of females are physics majors, as seen in Table 1. Females are less likely than males to be engineering majors, but about twice as likely as males to be other science or non-science majors. Over 80% of the students are White, about 10% are Asian, and about 8% are

African American, Hispanic, or Native American. There are only small differences in the distributions of ethnicity by gender. These frequencies are presented in Table 1.

Table 1. Frequencies for gender, student declared major, and ethnicity for all students in the study, that is, students who enrolled in introductory physics between Spring 2004 and Spring 2007.

Gender (N = 3728)	%		
Male	75.8		
Female	24.2		
Major (N = 3728)	%	% of males	% of females
Physics	5.5	5.6	5.2
Engineering	51.8	55.6	39.9
Other Science	18.1	14.1	30.5
Non-Science	9.9	8.4	14.6
Undeclared/Other	14.8	16.4	9.9
Ethnicity (N = 3514)	%	% of males	% of females
Asian	8.9	8.1	11.5
African American	1.3	1.3	1.1
Hispanic	6.1	6.1	6.3
Native American	0.8	0.9	0.7
White	81.1	81.9	78.7
Foreign	1.7	1.7	1.8

C. Performance and Background Measures

Of primary interest in this study is to what degree males and females differ on measures of background and preparation, and to what degree these differences contribute to the observed gender gap. Conceptual performance, as measured by the Force and Motion Concept Evaluation (FMCE) [31], serves as the focus of the study. The FMCE post-test score for each student is used as a measure of the student's conceptual knowledge of physics at the end of the semester. Only students with matched pre- and post-test data are included ($N=2099$). Additional evaluation of student performance in the course is captured by homework, exam, participation, and course grades, which were collected from the instructor in each course.

While we can look at a variety of measures to assess student performance (post-test,

normalized gain, course grades, etc.), we focus on the post-test as an objective measure of what students know at the end of the semester. While normalized gain can also be used, we opt to use the post-test as a measure of students' physics knowledge after a semester of instruction, rather than their gain in knowledge over the course of a semester. Several researchers have pointed out that there are biases associated with using normalized gain [32,33]. Course grades are subjective and measure more than just performance on a single instrument. Prior work has mostly focused on subjective measures of student performance and the self-reported backgrounds of students.

Data have been gathered (from university records [34]) on students' background knowledge and their preparation for college physics. Prior academic performance is captured by students' high school GPA, while the FMCE pre-test is used to measure students' prior conceptual understanding of physics. Four mathematics tests are combined to form a measure of students' prior knowledge of mathematics. The four tests include the SAT-Math, the ACT-Math, and two diagnostic exams that are given to students before their freshman year at CU. One diagnostic exam (denoted APPM test) is given through the Applied Mathematics department to students in the College of Engineering and Applied Science. The second exam (denoted ASMATH test) is given through the Mathematics department to students in the College of Arts and Sciences. Both diagnostic exams are used to help place students in the appropriate math course and do not count towards any course grade. Scores on each of the four tests were similarly correlated with the FMCE post-test ($0.3 < r < 0.4$), and were also highly correlated ($0.5 < r < 0.7$) with each other. To get a measure of prior math knowledge for almost every student and to avoid having multiple variables that contained the same information, the scores on the four tests were combined. The scores for each test were first normalized (converted to z -scores [35]). Each student's normalized scores were averaged to get a combined measure of mathematics

knowledge. Each student's combined math score is a composite of whichever of the four tests that the student took. Student course preparation for college physics is measured by how many years of high school physics and calculus a student had taken. Data were not available on the grade that students received in their high school courses.

In addition to students' prior content knowledge, data were also collected on their attitudes and beliefs about physics and about learning physics. Attitudes and beliefs are measured by the Colorado Learning Attitudes about Science Survey (CLASS) [36]. The CLASS questions are classified into eight categories of student beliefs. The survey is made up of 42 statements and students respond on a Likert-like scale. Each response is coded favorable, neutral, or unfavorable, based on whether the response agrees or disagrees with the expert response. Students are then given a % favorable and a % unfavorable score on each category. Favorable pre-test scores on each category are used as measures of students' incoming beliefs. Favorable post-test scores and shifts (*post - pre*) are used as measures of students' attitudes and beliefs at the end of the semester and to measure change in attitudes and beliefs, respectively.

We note that the several assessments used throughout the study only measure student *performance* on these instruments – however we use them as a proxy measurement of student understanding and actual attitudes and beliefs upon entry and exit. We recognize these instruments may be measuring more, such as test taking ability, and may differ by gender. In particular, McCullough [37] found that by changing the context of questions on the FCI to gender neutral or female contexts, male and female students responded differently. Other researchers identified differences in how students responded on the FCI when asked to mark the answer that they believed and the answer they thought scientists would give. Females answered differently in each case more often than males [38]. Still others have pointed out that the format

of questions that are typically asked in physics classes (multiple choice questions) may disadvantage females [39]. While these studies question the validity of these instruments, we note: a) we are using the standard measures that have been adopted by the community and b) we are analyzing *shifts* on these instruments, which allows us to normalize students against themselves.

The current study aims to identify several prior factors that influence student performance on a research-based mechanics conceptual learning instrument using data collected from university applications for students in an introductory, calculus-based physics course.

D. *Study Sample*

The FMCE is administered in the first and last weeks of class during recitation, and only those students who attend both weeks take the pre- and post-FMCE. As a result, we explore the possibility of sampling bias. Of the 3,728 students who took introductory physics during the semesters included in this study, 2,099 students (56%) took both the pre- and post-FMCE. Comparing the populations who did and did not take the FMCE, we find that females were more likely to take the FMCE than males: 63% of females versus 58% of males took the FMCE. The course grades (on a scale from 0.0 to 4.0) for males and females in each group are shown in Table 2. Not only are the average course grades of students who take the FMCE higher, but the gender gap in course grades for this group is smaller than for those who do not take the FMCE. By focusing on the FMCE as a measure of learning, we limit the sample of students included in the analysis and exclude primarily those with lower course grades. Also, the smaller gender gap in course grades among those who take the FMCE suggests that we may be underestimating the gender gap.

Table 2. Average course grades for males and females who did and did not take the FMCE. Course grades are on a 0 to 4.0 scale [40].

	Males			Females			Differences	
	N	Mean	SD	N	Mean	SD	M – F	p value
Students without FMCE	1152	2.14	1.2	315	1.89	1.1	0.25	0.001
Students with FMCE	1566	2.82	0.8	533	2.74	0.8	0.08	0.086

IV. Results

We present the results in three sections, corresponding to the three research questions that we seek to answer in this study: 1) Can we identify gender differences in the introductory physics course? 2) Are measures of student background correlated with student performance in Physics 1? and 3) Can differences in student background account for the gender gaps in student conceptual performance?

A. Identifying Differences By Gender

1. College Course Performance Differences

We are first interested in whether interactive engagement methods can reduce or eliminate the gender gap at our institution [41]. Figure 1 presents the average pre- and post-FMCE gender gaps for partially interactive and fully interactive courses [42]. The pre-test gender gaps are statistically equal for partially and fully interactive courses [$t(2098)=1.17, p=0.24$]. However, the post-test gender gap is significantly smaller for fully interactive courses than for partially interactive courses (14% versus 8%) [$t(2098)=2.27, p=0.02$], but it has not been eliminated [43]. Similar gender gaps are found when looking at normalized learning gains. For both partially and fully interactive courses males have a higher average normalized learning gain than females, although the difference between males and females is smaller for fully interactive courses.

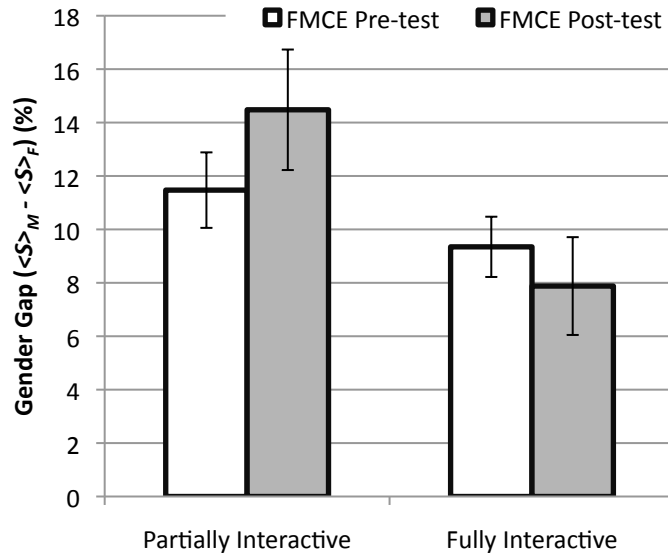


Figure 1. Pre- and post-test gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) for partially and fully interactive courses. Student performance on the FMCE is averaged over three semesters (partially interactive) and four semesters (fully interactive). Error bars represent the standard errors of the mean. There is no statistically significant shift in the gender gap for either partially or fully interactive courses.

Figure 2 presents the pre- and post-test gender gaps for each semester included in the present study. A different instructor lectured during each of the seven semesters. Counter to previous findings [14], we find that the size of the post-test gender gap is not independent of instructor. There is some consistency within the IE 1 and IE 2 courses. In all three IE 1 courses the gender gap increased (though not significantly) from pre- to post-test. In three of the four IE 2 courses, the gender gap decreased (though not significantly). But there is one IE 2 class (Semester E1-1) in which the gender gap increased. These findings suggest that the implementation of a fully interactive curriculum alone is not enough to eliminate, or even reduce the gender gap. It appears that the manner in which courses are implemented is significant [30]. Furthermore, the way in which the curriculum is enacted may appear to impact the gender gap, however we find (below) in these cases that differences in the gender gap from semester to semester can largely be accounted for by background differences of the students.

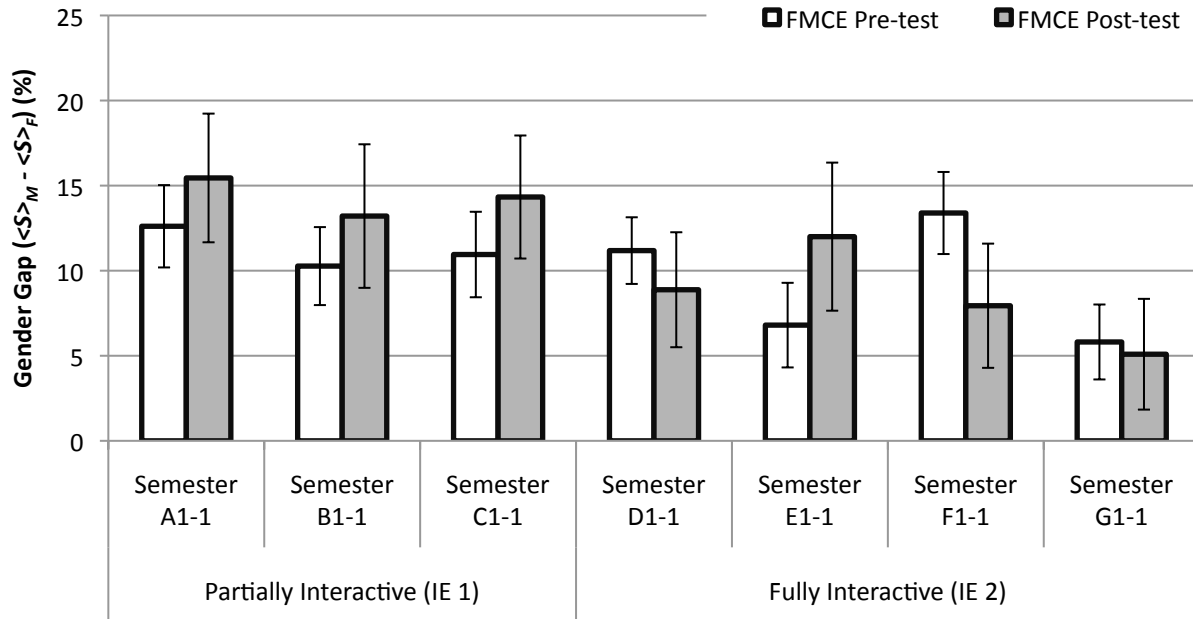


Figure 2. Gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) in each semester. The data represent seven different instructors and over 2000 students. Error bars represent the standard errors of the mean.

As reported above, females have lower normalized learning gains than males, meaning that females learn a smaller percentage of what they did not already know coming into the introductory course. In addition to looking at normalized learning gain, we also look at male and female average absolute gain ($G = post - pre$). The average absolute gain for both males and females is statistically significantly higher in IE 2 courses than in IE 1 courses [males: $t(1564) = 4.31, p < 0.01$; females: $t(531) = 4.33, p < 0.01$], but in neither pedagogical approach is the difference between average absolute gains for males and females significantly different [IE 1: $t(788) = 1.62, p = 0.11$; IE 2: $t(1307) = -0.96, p = 0.34$]. In IE 1 courses, the average gain of males and females is $G_M = 32\%$ and $G_F = 29\%$. In IE 2 courses, the average gain of males and females is $G_M = 37\%$ and $G_F = 39\%$. In three of the four IE 2 courses females had larger average absolute gains than males, while in all three IE 1 courses females had lower average absolute gains than males, though none of the differences was significant. Some have suggested that absolute gain

may be a more appropriate way to assess learning [44]. We observe that in terms of absolute learning gain, there is no statistically significant gender difference in any individual course or across all courses.

Course grades were examined to determine if males and females perform differently on course grades or any components of the course grades. For each of the seven semesters of the mechanics course males' and females' scores are averaged on homework, participation, exams, and total course grade. In all of the introductory courses exams make up 60% to 65% of the course grade, homework counts for 25% to 35%, and participation makes up the remainder. The difference between the average male and female's scores in each component ($\langle S \rangle_M - \langle S \rangle_F$) is calculated for each class. These differences for each semester, along with the average differences across all semesters, are shown in Table 3. For several courses the participation grade was included in the homework grade and could not be extracted.

Table 3. Analysis of students' course grades. Each column contains the difference between the average scores for males and females ($\langle S \rangle_M - \langle S \rangle_F$). Error (shown in parentheses) is computed from the standard errors of the mean for males and females added in quadrature. The * indicates the gender difference is statistically significant at the $p < 0.05$ level (via two-tailed t -test).

	Participation (%)	Homework (%)	Exams (%)	Course GPA (4 pt scale)
Semester A1-1		-5.0 (2.0) *	3.4 (1.2) *	0.11 (0.11)
Semester B1-1		-4.8 (1.8) *	3.7 (1.6) *	0.10 (0.11)
Semester C1-1		-5.0 (2.0) *	6.3 (1.5) *	0.10 (0.10)
Semester D1-1	-6.6 (1.6) *	-7.6 (1.9) *	4.9 (1.6) *	0.04 (0.10)
Semester E1-1	-4.9 (1.8) *	-2.9 (1.9)	5.2 (1.5) *	0.17 (0.11)
Semester F1-1	-8.1 (1.8) *	-2.0 (2.0)	4.8 (1.6) *	0.15 (0.12)
Semester G1-1	-3.0 (1.6)	-3.0 (2.0)	3.3 (1.4) *	0.06 (0.11)
Average	-5.6 (0.9) *	-4.5 (0.8) *	4.5 (0.6) *	0.11 (0.04) *

There was no significant gender difference in total course grade in any individual course of the seven semesters in the study. Males outscore females by about 5 percentage points on exams and females outscore males by about 5 percentage points each on homework and

participation. These differences offset one another and result in course grades that are not significantly different. Because of the consistent gender gap observed from semester to semester we find the difference in overall course grades of males and females is statistically significant, when averaging over all seven semesters.

In addition to looking at performance, we can also explore how the attitudes and beliefs of males and females change over the course of the semester, and whether there are any gender differences. Developers of the CLASS identified gender differences on almost half of the statements, and found that on average, females were less expert-like in their beliefs than males at the end of an introductory, calculus-based physics course [36]. Here, we present the average shifts (*post – pre*) for males and females overall and in each category for six semesters of the introductory, calculus-based physics course [45]. Shifts indicate how much students' attitudes and beliefs have changed from the beginning to the end of the semester. As Figure 3 shows, all of the shifts are negative, indicating that both males and females shift towards less expert-like beliefs about physics over the course of the introductory physics class [46]. In addition, females have *more* negative shifts than males overall and in each category. The difference in shifts is significant [$t(1897) > 2.23, p < 0.03$] for the three problem solving and two conceptual categories, and the gender difference in shifts are marginally significant overall [$t(1897) = 1.94, p = 0.05$] and in the real world category [$t(1897) = 1.85, p = 0.07$]. Females have pre-test scores that are similar to or lower than males' pre-test scores in each category (as shown in parentheses in Figure 3), so the larger negative shifts result in an increase of the gender gap in CLASS scores from pre- to post-test in all categories.

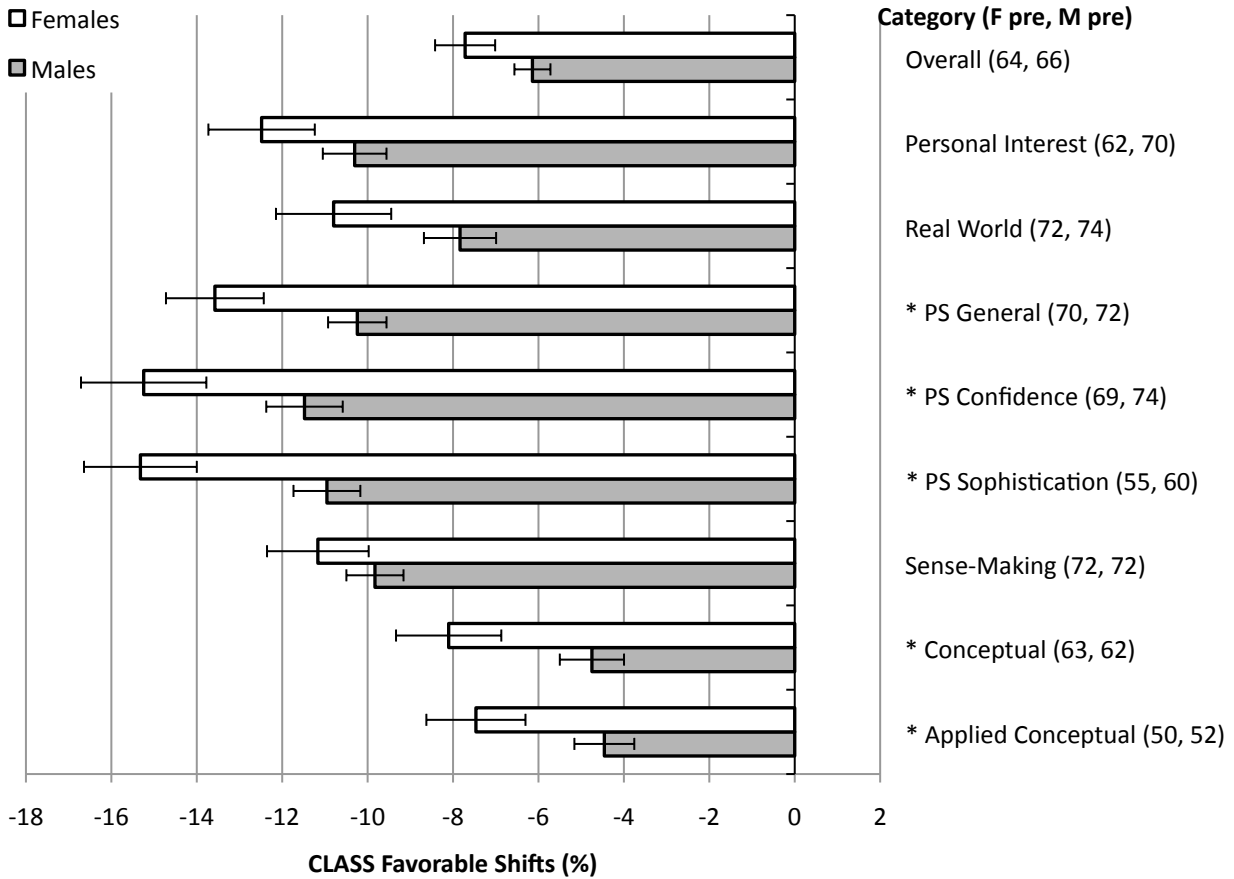


Figure 3. Average shifts (*post – pre*) for males and females on each of the CLASS categories. Note that all shifts are negative, meaning both male and female students shift toward less expert-like attitudes and beliefs about physics. The * indicates that the difference in shifts for males and females is significant ($p < 0.05$). Values in parentheses are female and male average pre-test scores. Females have more negative shifts in each category than males.

The same trends exist for the shifts in IE 1 courses and IE 2 courses separately. Females always have more negative average shifts than males in both pedagogical approaches. There are very few significant differences when comparing IE 1 to IE 2 courses for males and females separately. Aside from some small differences, partially and fully interactive courses have similar (negative) influences on students’ attitudes and beliefs about physics.

2. Academic Background Differences

In the previous section we reported observed differences between males’ and females’ performance during the introductory physics course. Here, we examine the background and

preparation of males and females. Male and female averages for each of the background variables collected are presented in Table 4. Note that not all data are available for all students, as is the case in any course. As a consequence of missing data the reported averages may be biased due to sampling error. We present them regardless as they are the best estimates we have of the values for all students who enroll in introductory physics.

Males significantly outperform females on both the SAT-Math and ACT-Math tests. Surprisingly, there are no gender differences on either of the CU diagnostic exams that are given at the beginning of freshman year. Because of the differences in SAT and ACT scores, there is a significant difference between the average combined math score of males and females.

Females have a higher average high school GPA than males by about 0.2 points. On average females take less high school physics than males, but they take about the same amount of high school calculus. These same data can be represented another way by looking at the percentage of males and females that have at least one year of high school physics and calculus. From Table 5, 89% of males and only 80% of females in introductory physics completed at least one year of high school physics. Only small percentages, 16% of males and 11% of females, took two years of high school physics. There are only minor differences in the fraction of males and females who take high school calculus; 67% of males and 70% of females took at least one year of high school calculus. It is interesting to note that both males and females are more likely to take high school physics than to take high school calculus.

Table 4. Male and female average values for all background variables that were collected. The range of possible scores for each variable is shown in parentheses. The effect size is calculated as $ES = (\langle S \rangle M - \langle S \rangle F) / SD$, where the SD for all students is used. Significant differences (via two-tailed t -test) exist between males and females on almost all of the background variables.

	Males			Females			Differences		
	N	Mean	SD	N	Mean	SD	M - F	Effect Size	p value
High school GPA (0.0 – 4.0)	2712	3.55	0.4	869	3.74	0.4	-0.19	0.47	<0.01
Years High School Physics	2463	1.08	0.6	758	0.95	0.5	0.13	0.24	<0.01
Years High School Calculus	2445	0.78	0.6	748	0.81	0.6	-0.03	0.05	0.21
SAT–Math (200 – 800)	1903	645	65	629	623	75	22	0.33	<0.01
ACT–Math (1 – 36)	2130	28.1	4	732	27.6	4	0.5	0.14	<0.01
APPM test (0 – 30)	1189	22.0	4	255	21.9	5	0.1	0.04	0.63
ASMA TH test (0 – 30)	349	17.1	2	128	17.3	2	-0.2	0.06	0.60
Math combined (z -score)	2744	0.024	0.9	869	-0.184	1	0.21	0.23	<0.01
CLASS pre-test (0 – 100)	1380	65.7	16	522	63.6	16	2.1	0.13	0.01
FMCE pre-test (0 – 100)	1566	32.2	21	533	22.0	16	10.2	0.49	<0.01
FMCE post-test (0 – 100)	1566	67.3	27	533	56.8	29	10.4	0.37	<0.01

Table 5. Percentages of males and females that have taken high school courses in calculus and physics. The * indicates that the difference in percentages is significant via χ^2 test, $p < 0.01$.

	% of Males	% of Females
1 yr. HS Physics*	88.7	79.7
2 yrs. HS Physics*	15.5	10.7
1 yr. HS Calculus	67.3	69.9

B. *Correlation of Student Background with Student Conceptual Performance*

We have identified several aspects of the introductory physics course in which gender differences exist: conceptual surveys, course grades, attitudes and beliefs, and student background and preparation. The next step is to determine which, if any, of the student background factors are correlated with student performance on the conceptual survey, and could therefore be contributing to the gender gap in observed post-test scores.

We first ask: do differences in male and female average post-test scores exist when students are grouped according to their pre-test score? If males and females with similar pre-test scores have different post-test scores, then there would be evidence that despite equal performance on measures of background physics knowledge, there is differential learning by gender. Students are binned by their FMCE pre-test score (each bin contains about equal numbers of students, $N \sim 420$), and then the average FMCE post-test score is calculated for males and females in each bin. The results are plotted in Figure 4. The same trends that are described below exist for a range of reasonable bin sizes.

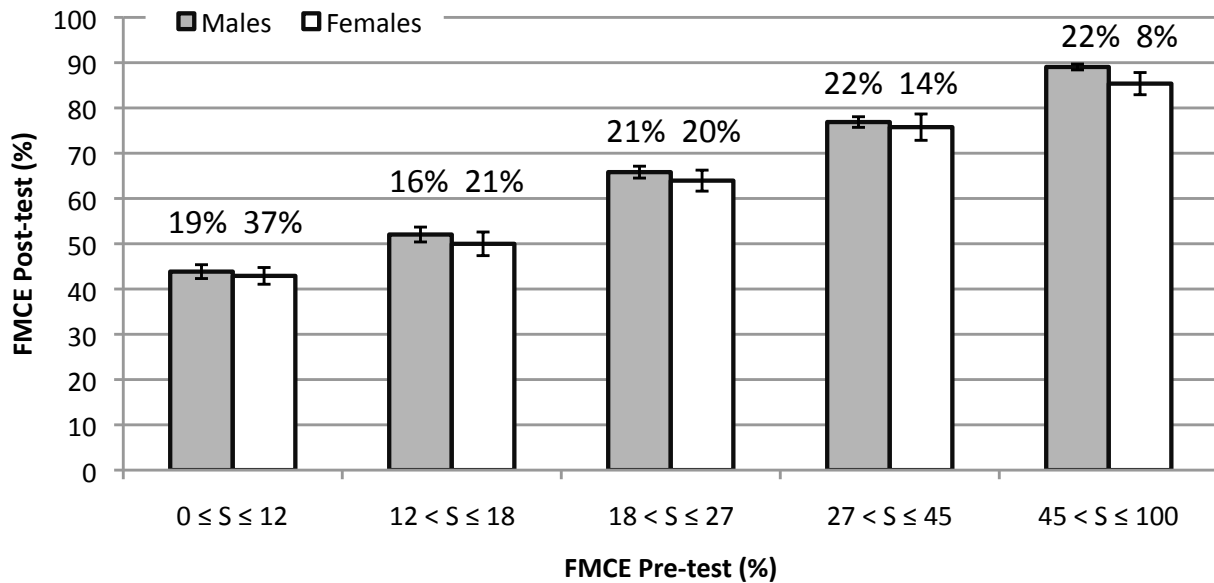


Figure 4. Average FMCE post-test scores for females and males with matched FMCE pre-test scores. The percentages above each bar represent the percentage of the males (or females) from the total in each bin. The error bars represent the standard error on the mean. There are no significant differences between males and females in any individual bin.

Students who have similar pre-test scores have similar post-test scores, regardless of gender. There were no statistically significant differences [t 's > 1.87, p 's > 0.15] in any individual bin, i.e. between males and females who scored similarly on the pre-test. Though the differences in each bin are not significant, males consistently score higher than females in all bins. We also see a correlation ($r=0.56$) between FMCE pre- and post-test scores. These same trends exist for each individual semester.

Though a similar number of students is in each bin, a higher percentage of the females fall into the low pre-test bins. The percentages above each bar in the plot (Figure 4) represent the percent of the males, or females, who fall into that bin. 58% of females versus 35% of males fall into the lowest two pre-test bins, while 22% of females versus 44% of males fall into the highest two pre-test bins. Thus, a dominant source of the observed gender difference (from Figure 2) is

attributable to the low pre-test scores of females combined with the correlation between pre- and post-test scores.

The same trend exists for normalized learning gain and absolute learning gain; students with similar pre-test scores have similar normalized and absolute gains, regardless of gender. We also see a correlation, albeit weaker ($r=0.3$), between FMCE pre-test score and normalized gain, and a correlation ($r=-0.2$) between FMCE pre-test score and absolute gain. Furthermore, the results look the same whether students are in IE 1 or IE 2 courses.

To determine if taking high school physics influences the pre- and post-FMCE scores, we examine males and females who did and did not have a high school physics course. The results are presented in Table 6. Looking first at the pre-test, students who had taken high school physics score significantly higher than students who did not have high school physics [males: $t(1412)=5.48, p<0.01$; females: $t(460)=4.69, p<0.01$]. The difference is greater for males than for females. The gender gap for students who took high school physics is about 10 percentage points, while the gender gap for students without any high school physics is only about 4 percentage points. The gender gap on the post-test for those students who take high school physics is 9 percentage points, statistically the same as the pre-test gap. But, for students who had no high school physics, the gender gap on the post-test is 16 percentage points (significantly larger than the pre-test gap). Similar to the pre-test, those students who took high school physics had higher average post-test scores than those students who did not [males: $t(1412)=2.12, p=0.03$; females: $t(460)=4.51, p<0.01$]. But, the gap is larger for females than for males.

Table 6. Average FMCE pre-test scores for males and females who did and did not take high school physics. The * indicates that the difference is statistically significant at $p < 0.01$.

	FMCE Pre-test			FMCE Post-test		
	Males	Females	M – F	Males	Females	M – F
Had HS Physics	33.5	23.9	9.6 *	68.0	58.9	9.1 *
No HS Physics	20.2	15.8	4.4 *	60.7	44.9	15.8 *
Phys. – No Phys.	13.3 *	8.1 *		7.3 *	14 *	

C. Estimation of the Impact of Student Background on the Gender Gap

1. Multiple Regression Analysis

Having identified several background variables that are correlated both with gender and with student performance on the FMCE, the next step is to model the post-test scores using multiple regression [47]. However, due to the nature of our data, we cannot strictly interpret the statistical significance of the results, as they are likely to be biased. Because of ceiling effects, non-normal data, heteroskedasticity, and non-random sampling [48], our data do not meet the strict assumptions of multiple regression that allow for unbiased interpretation of statistical significance. We can, however, use the regression analysis to describe the patterns in our data, without needing to meet the assumptions of multiple regression [49]. The results of the multiple regression analysis will describe the relationship between a student’s post-test score and the values of several background variables for that student. Using this relationship, we estimate the difference in post-test scores for a male and female with all background variables being equal. In this way, we will determine how much of the gender gap can be accounted for by factors other than gender.

The post-test scores are modeled according to the equation,

$$FMCEPOST = b_0 + b_1 \times FEMALE + \sum_{k=2}^N b_k \times VAR_k$$

where $FMCEPOST$ is the post-test score on the FMCE, $FEMALE$ is a dummy variable that is 1 for females and 0 for males, and VAR_k are the other background variables that are included in the model and any cross terms between $FEMALE$ and the other background variables. The b_k are the coefficients for each term, and the multiple regression analysis gives estimates for these coefficients. The coefficient of the $FEMALE$ variable (b_1) gives the difference between a male's and a female's scores, all other factors being equal. It is this coefficient that we are ultimately interested in.

We are modeling students' FMCE post-test scores rather than their absolute or normalized gain, because we are primarily interested in reducing the gender gap in post-test scores. By modeling the post-test, we can determine what factors influence the post-test score, and could therefore contribute to the gender gap. Each of the possible confounding variables is included in the regression analysis. Variables are entered sequentially in order to find the parsimonious combination of factors that best predicts the post-test score for each student. The best model will be judged based on the size of the coefficients, the increase in Multiple R^2 (the fraction of variation in post-test scores that is accounted for by the variables in the model), and to a lesser degree, the significance of variable coefficients (although as mentioned above, the p values may be biased).

As stated above, not all data were available for all students. This being the case, only a subsample of the students who took the introductory course was used in the multiple regression analysis. Recall that only 2,099 of the 3,728 students who enrolled in introductory physics between Spring 2004 and Spring 2007 took the FMCE pre- and post-tests. Of these 2,099 students, complete data [50] were available for 1,027 students. These 1,027 students make up the sample used in the analysis. It is important to keep in mind that the sample used in this analysis

is not representative of all students who enroll in introductory physics. The percentage of females in this sample is 29%, which is higher than the 24% for the population. It appears that females are more likely to take voluntary surveys (such as the FMCE and CLASS) which results in a slight oversampling of females. Also, the average course grades of females and males are higher for students in this sample than for students not in the sample. We again point out that by looking only at this sample of students we may be underestimating the gender gap. Furthermore, the results that we report below apply only to students in the sample, and cannot be extrapolated to describe students not in the sample.

The results of the multiple regression analysis are shown in Table 7. Four models are reported, starting with a bivariate model that includes only gender, and then additional variables are added in each successive model. The table contains the coefficient estimates (b_k) for each model as well as the model level statistics. The variables entered in each successive model are not only significant, but they increase R^2 substantially (the additional variance explained by each model is significant via F -test at the $p < 0.01$ level). The R^2 for the final model is 0.44, such that the variation in the independent variables explains 44% of the variation in post-test scores.

We are interested in the difference between males' and females' post-test scores after controlling for several prior factors. In Model 1, where only *FEMALE* is included as an independent variable, the gender difference is 10.7 points. This is just the average difference in post-test scores between males and females in this sample. In Model 2, several covariates that are correlated with the post-test are added. When previous physics knowledge (FMCE Pre-test), previous math knowledge (Combined Math Score), and previous attitudes and beliefs (CLASS Pre-test) are controlled, the gender difference drops to 4.3 points. Already, there is a substantial

reduction in the gender difference once previous physics and math knowledge and attitudes and beliefs are accounted for.

Table 7. Coefficient estimates and model level statistics for each multiple regression model.

	Model 1	Model 2	Model 3	Model 4
Model-level Statistics				
Multiple R^2	0.03	0.42	0.42	0.44
F statistic p -value	<0.001	<0.001	<0.001	<0.001
Residual Standard Error	27.3	21.2	21.1	20.9
Predictors				
	b_k	b_k	b_k	b_k
Intercept	67.2	29.8	31.5	32.9
Female	-10.7	-4.3	-10.0	-9.2
FMCE Pre-test		0.63	0.59	0.59
Combined Math Score		7.4	7.2	7.2
CLASS Pre-test		0.25	0.24	0.26
Semester A1-1 (IE 1)				1.3
Semester B1-1 (IE 1)				-5.6
Semester C1-1 (IE 1)				-8.7
Semester E1-1 (IE 2)				-2.9
Semester F1-1 (IE 2)				-0.93
Female \times FMCE Pre-test			0.23	0.20

When regressing post-test on pre-test for males and females separately, we observe that the two regression lines have different slopes. Model 3 includes an interaction term that allows the slope of the FMCE pre-test variable to differ for males and females. This term is the product of the two variables *FEMALE* and *FMCEPRE*. Since *FEMALE* is 0 for males and 1 for females, the interaction term is 0 for all males and is equal to the pre-test score for all females. The inclusion of the interaction term in the model suggests that the pre-test score differently predicts the post-test score for males and females. The interaction term also needs to be taken into account when estimating the gender difference. The gender difference now depends on the pre-test score.

To get a final estimate of the gender difference, we turn to Model 4. In this model, variables are added to take into account the semester that students took introductory physics. Controlling for semester is important for two reasons. First, by including a variable that controls for the semester that they took physics, some dependence among students due to taking physics at the same time is eliminated. Second, the average post-test scores are different in each semester. Including a semester variable will account for any differences that happen by semester which contribute to the post-test scores. Although we have no further information about specific aspects of each semester that could contribute to the differences, by including the semester variables we can see if there are differences once other prior factors are accounted for. The base case in Model 4 is semester G1-1 (meaning there is no variable included for this semester). This means that the coefficients of each semester variable give the average difference between semester G1-1 and that semester, after all other variables have been accounted for. For example, controlling for pre-test, math knowledge, and attitudes and beliefs, the average difference between semester G1-1 and semester C1-1 is about -8.7 points. Note that some of the differences are substantial, suggesting that even though the courses look similar according to the curriculum, how the curriculum is enacted may differentially influence whether students learn [30].

With Model 4, a final estimate of the difference between a male's and a female's post-test scores, controlling for several other factors, can be estimated. This difference is given by

$$M - F = 9.2 - 0.2 \times FMCEPRE .$$

The average pre-test score for this sample is 30.3. The gender difference for a male and a female with the average pre-test, and all other variables equal, is 3.2 points. This is a substantial reduction from the 10.7 point difference that is observed just by subtracting the average male and

female post-test scores. Controlling for student background, we account for 70% of the observed gender gap. The effect size,

$$ES = \frac{\langle post \rangle_M - \langle post \rangle_F}{SD_{post}},$$

where SD_{post} is the standard deviation for all students, went from 0.39, when no background variables were controlled, to 0.11, when measures of student background are controlled.

The resulting expression for the gender gap predicts that for males and females with pre-test scores above about 45%, the gender gap reverses sign. Females with pre-test scores greater than 45% are predicted to have *higher* post-test scores than males with the same pre-test. While this result is encouraging, we need to be cautious. There are very few data for students with pre-test scores above 45%, especially for females. Only 8% of the females and 20% of the males in the sample have pre-test scores greater than 45%. Because there are not many data for students with higher pre-tests, we cannot be sure that the predictions made in this region are accurate. The same can be said for very low pre-test scores. Only 10% of the students have pre-tests lower than 10%.

Excluded Variables

While the final model includes many variables that one might suspect would influence post-test scores, there are several variables that are not included. All of the variables listed above (Table 1 and Table 4) were included in the analysis, but none were found to contribute significantly to the model beyond those variables already included in Model 4. Looking more closely at some of the variables that are not included in the final model offers additional information.

Years of high school physics was somewhat correlated with the post-test ($r=0.2$), but was also correlated with the pre-test ($r=0.3$). For this reason, we suspect that years of high school

physics and the pre-test were contributing some of the same information about the post-test score. Because the pre-test was more highly correlated with the post-test ($r=0.6$), we chose to include that in the model over years of high school physics. In addition, others have pointed out that the specifics of the high school physics class are important [22,23,24], and that information may have been more useful in the model than just years of high school physics. Similar conclusions were drawn about years of high school calculus and the combined math score. There is not very much variation in high school GPA ($M=3.7$, $SD=0.3$), as only students who were admitted to CU and who took an introductory physics course are included. We suspect that the lack of variation in GPA and its low correlation with post-test score ($r=0.1$) made it less likely to be a useful predictor of post-test score.

Students' declared major also was not a significant predictor of post-test score. This suggests that after accounting for background differences, there is no difference in the post-test scores of students by major. We also found that ethnicity was not a significant predictor. We suspect that this is largely due to the small numbers of minority students who enroll in the physics course at CU, which makes it difficult to accurately estimate the influence of ethnicity on post-test score.

There were also several interaction terms that we attempted to include in the model. Notably, we included an interaction between the variable *FEMALE* and each semester variable. None of these interactions was significant, meaning that the gender gap was the same in each semester after controlling for previous knowledge and attitudes. This suggests that differences in the post-test gender gap from semester to semester (Figure 2) can be accounted for by differences in students' previous knowledge and attitudes. Interaction terms between gender and combined math score and between gender and CLASS score were also included, but neither

contributed to the post-test score. This suggests that math score and CLASS score equally impact the post-test score for males and females.

2. Logistic Regression Analysis

The previous multiple regression analysis gave us only a description of the data. Another way to analyze the data, which will allow for interpretation of the statistical significance of the results, is to use logistic regression analysis [51]. Logistic regression is used when the outcome variable of interest is a categorical variable rather than a continuous variable, for example, passing or failing rather than a raw score. While using this method allows us to make statistical claims, we lose the ability to predict students' actual post-test score and can only predict whether they will score above some threshold. What is gained in statistical specificity is lost in richness of the data analyzed. To model the data using logistic regression, the FMCE post-test variable is converted into a categorical variable (with any reasonable number of categories). The analysis was run for several threshold values (20%, 40%, 60%, 80%) and also for several numbers of post-test categories (2, 3, 4, 5). The results were similar for all threshold levels and number of categories, so we present the results for a threshold of 60% and two post-test categories here.

The frequencies of males and females who score above and below 60% on the FMCE post-test are presented in Table 8. Note that we are using the same sample of 1,027 students that was used in the multiple regression analysis above. We observe that 64% of the males and 49% of the females scored above 60% on the post-test. This difference in percentages is significant ($\chi^2(1, N=1027)=22.03, p<0.01$). Males and females are not equally likely to score above 60%. A gender gap is present in this new measure of student physics knowledge at the end of the course. This difference is the gender gap that we are concerned with in the logistic regression analysis.

Table 8. Percentages of males and females who score above and below 60% on the FMCE post-test. These are the percentages for the sample of 1,027 used in the logistic regression analysis. The difference in percentage of males and females that score above and below 60% is significant ($p < 0.01$).

	Males	Females
FMCE post-test > 60%	64.3%	48.5%
FMCE post-test < 60%	35.7%	51.5%

In logistic regression, rather than modeling the raw dependent variable, the *logarithmic odds* of the dependent variable is modeled. In this context, odds is defined [51] as the probability of an event occurring divided by the probability of an event not occurring. The post-test data are modeled according to the equation,

$$\ln(\text{odds}(\text{FMCEPOST} > 60\%)) = b_0 + b_1 \times \text{FEMALE} + \sum_{k=2}^N b_k \times \text{VAR}_k .$$

Given that the gender gap in this analysis is the difference in odds of scoring above 60% for males and females, we are interested in whether the difference in odds can be explained by factors other than gender. To determine the difference in odds, we are again interested in the coefficient of the *FEMALE* variable, b_1 . The odds for a male and a female, all other variables being equal, are related according to the equation,

$$\text{odds}_F = e^{b_1} \times \text{odds}_M .$$

The logistic regression analysis estimates the coefficients of each variable (as in the multiple regression analysis), which then allows a prediction of each student's odds of scoring above 60%. In each model we are interested in 1) whether the coefficient of *FEMALE* is significantly different from zero (as indicated by the p value) and 2) whether e^{b_1} is less than, greater than, or equal to 1. The results of the logistic regression analysis are shown in Table 9. For each model the coefficient estimates (b_k) and p values are given, as well as an evaluation of e^{b_k} . Only gender is included in Model 1. In this model, the coefficient of *FEMALE* is significantly different than

zero ($p < 0.01$), and we see that the odds of a female scoring above 60% are about half the odds of a male scoring above 60%. Just as in multiple regression, when only gender is included in the model, the predicted difference between males and females is just the observed difference.

In Model 2, covariates of the post-test are controlled for, including prior physics and math knowledge and prior attitudes and beliefs. In this model, the coefficient of *FEMALE* is not significant ($p > 0.1$), meaning the odds of scoring above 60% for a female are not statistically different than the odds for a male, holding all other variables constant. A male and a female with the same background (as measured by the prior factors included in the model) are equally likely to score above 60% on the post-test.

The interaction term between the pre-test and gender was included in Model 3 [52]. Unlike the results using multiple regression, here there is no significant interaction between prior knowledge and gender ($p = 0.45$). The pre-test has the same effect on whether students score above 60% on the post-test for both males and females. Because the interaction term is not significant, we do not include it in Model 4.

Table 9. Logistic regression analysis results. In each model covariates are included to control for differences in student background. Gender is not significant in the final model.

	Model 1		Model 2		Model 3		Model 4	
Model-level Statistics								
Pseudo R^2 (Nagelkerke's)	0.03	0.43	0.43	0.43	0.43	0.43	0.45	0.45
Likelihood ratio p value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Predictors	b_k	e^b	b_k	e^b	b_k	e^b	b_k	e^b
Intercept	0.59	<0.01	-3.0	<0.01	-2.9	<0.01	-2.9	<0.01
<i>FEMALE</i>	-0.65	<0.01	-0.26	0.13	-0.53	0.18	-0.23	0.19
FMCE Pre-test			0.08	<0.01	0.08	<0.01	0.08	<0.01
Combined Math Score			0.56	<0.01	0.55	<0.01	0.55	<0.01
CLASS Pre-test			0.02	<0.01	0.02	<0.01	0.02	<0.01
Semester A1-1 (IE 1)							0.14	0.59
Semester B1-1 (IE 1)							-0.68	0.01
Semester C1-1 (IE 1)							-0.75	0.01
Semester E1-1 (IE 2)							-0.10	0.71
Semester F1-1 (IE 2)							-0.06	0.81
Female \times FMCE Pre-test					0.01	0.45		1.0

Again, the semester variables are included in Model 4. Note that including the semester variables doesn't have a substantial impact on the coefficient of *FEMALE*, it remains insignificant, but it does allow us to compare odds of students across semesters. The base case is again semester G1-1. Just as with the multiple regression analysis, there are some semesters (semesters B1-1 and C1-1) in which the odds of scoring above 60% are significantly different from the odds in semester G1-1. This regression analysis only allows for statistical comparison between semester G1-1 and the other semesters. Repeating the logistic regression analysis with all of the other semesters as base cases, we find that the odds in semesters B1-1 and C1-1 are statistically equal to one another, but statistically different from the other four semesters. The odds in semesters A1-1, E1-1, F1-1, and G1-1 are all statistically equal once prior factors are accounted for. Again, even though prior factors are accounted for, there are still statistically significant differences between some semesters.

With semester differences accounted for, the final estimate of the relationship between the odds for a female and the odds for a male, holding all other variables constant, is

$$odds_F = 0.8 \times odds_M.$$

This is smaller than 1 (but not statistically different from 1), meaning that the odds for males and females are statistically equal. By accounting for student background, the factor relating the odds of males and females has gone from 0.5 to 0.8. Using logistic regression and controlling for student background, we account for 60% of the observed gender gap in odds. The gender gap in odds can be largely accounted for by prior physics and math knowledge and prior attitudes and beliefs.

Limitations of the Study

The variance explained by the final models is about 45% of the variance in post-test scores [53]. There may be other prior factors that we have overlooked that could be important in helping to explain the post-test score and could contribute to the gender gap. Notably, we have not included any variables that characterize students' motivations, study or learning habits, or their reasons for being in the class. Socioeconomic status is a demographic variable that was not included in the model. A proxy for socioeconomic status (financial aid information) was available, but only for those students who applied for need-based financial aid, which was too limited a sample to include in the analysis. There are also other aspects of students' background that were not included (other high school courses, grades in high school courses, other components of standardized tests, etc.). As mentioned above, there are specific aspects about how a faculty member implements the curriculum that have not been accounted for. Only an overall semester variable was included, which does not contain more detailed information about how the curriculum was implemented.

There are limitations in the applicability of these regression results to the entire population of students in the study. The sample of students used in the regression analyses is only about 30% of the students that enrolled in introductory physics during the semesters included in the study. We reiterate that although the students in the sample are different from the population of all students, by using this sample of mostly high performing students (in terms of course grades) and given the larger gender gap in course grades among the students not in the sample, it is possible that we are underestimating the gender gap of all students.

Finally, there are potential limitations due to the reliability of the instruments that were used to assess learning and prior knowledge. While our attempts to examine shifts in student

performance allow us to normalize students against themselves, the broader scale concerns about gender-based biases, such as stereotype threat [54], still remain. Some hint of test-taking being a factor that differentially impacts female performance is the data on student grades. Consistently the males outperform the females on exams while females outperform males on homework and other course components.

V. Discussion and Conclusions

While the differential performance of male and female students is now well documented, the sources of the gender gap and routes to addressing this disparity have been less well understood. By examining the performance and background of nearly four thousand students who took introductory physics at the University of Colorado, we begin to understand the sources of and possible solutions to this challenge. Our present studies find that the gender gap exists well beyond measures of student conceptual learning. Student grades vary by gender, both in overall scores and by course component. We observe that males and females have different shifts over the course of the semester in their attitudes and beliefs about physics, suggesting that males and females are experiencing the same course in different ways. The physics and mathematics background and preparation of students coming into our courses also varies by gender.

In taking a closer look at the gender gap in measures of conceptual performance, we observe that the pre- and post-test gender gaps are not consistent from semester to semester. Although the regression analysis suggests that these differences in the *gender gap* from semester to semester can be accounted for by background, differences in the average post-test score of all students from semester to semester are present even after controlling for student background. Given that there is relative consistency on the large scale in these courses, it appears that instructor differences, the course specifics, the way in which the curricula are implemented, and

potentially, the fine-grained choices that are made with regard to content and course structure impact the overall performance of all students. While we observe differences in males' and females' post-test scores and in their normalized gains, we find no significant differences in average absolute gain on these measures of conceptual learning in any semester or overall. If learning is defined by absolute gain, rather than normalized to prior knowledge, there is no gender gap.

Several of the background measures correlate with student performance on the FMCE post-test, suggesting that part of the gender gap may be attributed to differences in student background. In particular, when we bin students by pre-test score, we find no difference in post-test scores between males and females with similar pre-test scores. This is not the case when only taking into account whether or not students took high school physics. The gender gap in post-test scores is present both between males and females who did take high school physics and those who did not. Furthermore, the gender gap is exacerbated for those students who did not take high school physics. While controlling for whether students take high school physics does not account for the observed gender gap, our data suggest that differences in students' pre-test scores and other measures of student background may account for a substantial fraction of the gender gap.

Both the multiple regression and the logistic regression models confirm this interpretation, showing that a majority of the gender gap can be accounted for by factors other than gender, explicitly. From the multiple regression analysis we find that only 3 points of the 11 point gender gap can not be accounted for by background factors. From the logistic regression analysis we find that the odds of a male and a female scoring above 60% on the post-test are not statistically different once background factors are accounted for. Taken together, the results of

these models suggest that the persistence of the gender gap is due in large part to differences in males' and females' preparation and background coming into the introductory course, and not explicitly due to their gender.

In one sense, it may be interpreted that gender does not play a role in measures of student achievement – the variation in FMCE post-test score may be attributed to other variables, notably pre-test score, student beliefs, and math achievement. For a given semester, male and female students make statistically indistinguishable grades. Such a stance would suggest that there is no explicit gender bias in the classes observed. Both males and females show learning gains from pre- to post-test. Nonetheless, in these classes we observe a gap in performance by gender and observe instances where over the course of instruction, this gap is increased.

Another interpretation is that of implicit bias – that is, those components of a class that are most heavily weighted and essential for success disproportionately favor male students. While course grades are statistically neutral overall for a given semester, male students are more likely to score higher on exams (which are weighted more heavily in a typical class). Further, over all semesters we find a small but significant difference in overall course grades. While the classes studied are introductory courses with no expectation of prior knowledge of physics, those students who arrive to the class with greater background knowledge (higher pre-test scores) are more likely to achieve high post-test scores and greater normalized learning gains. The class favors those students with stronger physics and math backgrounds – in this case, male students.

Such an arrangement of a class (or any social environment) plays to certain student backgrounds and when those backgrounds are correlated with particular demographic groups, it demonstrates bias. That is not to say this is an explicit or purposeful bias, but rather one that is the codified structure of systemic cultural bias [55]. Tatum refers to this as a “smog of bias” [56]

and others to the privileged preparation of some group (at the expense of others) as an “accumulated advantage” [57]. Recognizing that student preparation in physics or mathematics is a means by which this bias is propagated allows us as researchers and educators to proactively address the challenges of the gender gap in physics. Simply enacting research based reforms, or supporting current practices (the *status quo*) may improve aggregate student learning gains, but may also be promulgating the disparity of performance and lack of equity in our educational system.

Peer-reviewed publications based on this work

- L. E. Kost, S. J. Pollock and N. D. Finkelstein, Characterizing the gender gap in introductory physics. *Physical Review Special Topics – Physics Education Research* **5**(1), 010101 (2009).
- L. E. Kost, S. J. Pollock and N. D. Finkelstein, “The Persistence of the Gender Gap in Introductory Physics” in *Proceedings of the 2008 Physics Education Research Conference*, edited by C. Henderson, M. Sabella and L. Hsu (AIP Conference Proceedings, New York, 2008) p. 139.
- L. E. Kost, S. J. Pollock and N. D. Finkelstein, “Investigating the Source of the Gender Gap in Introductory Physics” in *Proceedings of the 2007 Physics Education Research Conference*, edited by L. Hsu, C. Henderson and L. McCullough (AIP Conference Proceedings, New York, 2007) p. 136.
- S. J. Pollock, N. D. Finkelstein and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physical Review Special Topics – Physics Education Research* **3**, 010107 (2007).

References (Chapter 2)

- [1] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *American Journal of Physics* **66**, 64 (1998).
- [2] E. F. Redish, *Teaching Physics With the Physics Suite* (John Wiley & Sons, Inc., Hoboken, NJ, 2003).
- [3] L. C. McDermott and E. F. Redish, Resource Letter: PER-1: Physics Education Research, *American Journal of Physics* **67**, 755 (1999).
- [4] E. Mazur, *Peer Instruction: A Users Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- [5] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- [6] P. Laws, P. Rosborough and R. Poodry, Women's responses to an activity-based introductory physics program, *American Journal of Physics* **67**, S32, (1999).
- [7] M. Schneider, Encouragement of women physics majors at Grinnell college: A case study, *The Physics Teacher* **39**, 280 (2001).
- [8] *Teaching the Majority: Breaking the Gender Barrier in Science, Mathematics, and Engineering*, edited by S. V. Rosser (Teacher's College Press, New York, 1995).
- [9] National Research Council, *To Recruit and Advance: Women Students and Faculty in Science and Engineering* (National Academies Press, Washington DC, 2006).
- [10] D. Halpern, J. Aronson, N. Reimer, S. Simpkins, J. Star and K. Wentzel, *Encouraging Girls in Math and Science* (NCER 2007-2003). Washington DC: National Center for Education Research, Institute of Education Sciences, U. S. Department of Education (2007). Retrieved from <http://ncer.ed.gov>.
- [11] National Science Foundation, *New Formulas for America's Workforce: Girls in Science and Engineering*, NSF 03-207 (Arlington, VA, 2003). Available from <http://www.nsf.gov/pubs/2003/nsf03207/start.htm> and *New Formulas for America's Workforce 2: Girls in Science and Engineering*, NSF 06-60 (Arlington, VA, 2006). Available from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf0660.
- [12] American Physical Society, Committee on the Status of Women in Physics, *Gender Equity Report* (unpublished, 2007). Available from <http://www.aps.org/programs/women/workshops/gender-equity/>.
- [13] R. R. Hake, "Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pre-test scores on mathematics and spatial

- visualization” in *Proceedings of the 2002 Physics Education Research Conference*, edited by S. Franklin, K. Cummings and J. Marx (AIP Conference Proceedings, New York, 2003).
- [14] M. Lorenzo, C. Crouch and E. Mazur, Reducing the gender gap in the physics classroom, *American Journal of Physics* **74**, 118 (2006).
- [15] D. Hestenes, M. Wells and G. Swackhamer, Force concept inventory, *The Physics Teacher* **30**, 141 (1992).
- [16] P. Heller, R. Keith and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *American Journal of Physics* **60**, 627 (1992).
- [17] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *American Journal of Physics* **60**, 637 (1992).
- [18] A. M. L. Cavallo, M. Rozman, and W. H. Potter, Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors, *School Science and Mathematics* **104**, 288 (2004).
- [19] J. Docktor and K. Heller, “Gender differences in both Force Concept Inventory and introductory physics performance” in *Proceedings of the 2008 Physics Education Research Conference*, edited by C. Henderson, M. Sabella and L. Hsu (AIP Conference Proceedings, New York, 2008) p. 15.
- [20] E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O’Brien, I. Rodriguez, and P. Pamelá, Toward equity through participation in Modeling Instruction in introductory university physics, *Physical Review Special Topics – Physics Education Research* **6**, 010106 (2010).
- [21] D. Hestenes, Toward a modeling theory of instruction, *American Journal of Physics* **55**, 440 (1987).
- [22] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: the influence of high school physics preparation and affective factors, *Science Education* **91**, 847 (2007).
- [23] P. M. Sadler and R. H. Tai, Success in introductory college physics: the role of high school preparation, *Science Education* **85**, 111 (2001).
- [24] R. H. Tai, and P. M. Sadler, Gender differences in introductory undergraduate physics performance: university physics versus college physics in the USA, *International Journal of Science Education* **23**, 1017 (2001).

- [25] D. E. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible “hidden variable” in diagnostic pre-test scores, *American Journal of Physics* **70**, 1259 (2002).
- [26] In general, two professors are assigned to each of the large-enrollment introductory physics courses at CU. One professor (the lead) is responsible for the lecture part of the course and the other professor (the back-up) is responsible for the recitation part of the course. The back-up instructor usually does not have much, if any, interaction with students, while the lead instructor is who the students see in lecture. This being the case, we will only talk about the role of the lead instructor in these courses.
- [27] CAPA, <http://www.lon-capa.org/>; Mastering Physics, <http://www.masteringphysics.com/>.
- [28] V. Otero, N. D. Finkelstein, R. McCray and S. Pollock, Who is responsible for preparing science teachers? *Science* **313**, 445 (2006).
- [29] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: implementing tutorials in introductory physics, *Physical Review Special Topics - Physics Education Research* **1**, 010101 (2005).
- [30] C. Turpen and N. D. Finkelstein, Not all interactive engagement is the same: Variations in physics professors’ implementation of *Peer Instruction*, *Physical Review Special Topics – Physics Education Research* **5**, 020101 (2009).
- [31] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton’s laws: The force and motion conceptual evaluation, *American Journal of Physics* **66**, 338 (1998).
- [32] E. Brogt, D. Sabers, E. E. Prather, G. L. Deming, B. Hufnagel and T. F. Slater, Analysis of the astronomy diagnostic test, *Astronomy Education Review* **6**, 25 (2007).
- [33] S. D. Willoughby and A. Metz, Exploring gender differences with different gain calculations in astronomy and biology, *American Journal of Physics* **77**, 651 (2009).
- [34] Data on students’ high school experience (high school courses, high school GPA, SAT-Math, and ACT-Math scores), student demographic data (gender, ethnicity, and declared major), and course data (semester of enrollment, course grade) were collected from the Office of Planning, Budget, and Analysis at CU Boulder. None of these data were self-reported.
- [35] For example, a student’s SAT-Math score, S , was converted to a z-score using the transformation, $z_S = (S - \langle S \rangle) / SD$, where $\langle S \rangle$ is the average SAT-Math score for all students and SD is the standard deviation of the SAT-Math scores.

- [36] W. K. Adams, K. K. Perkins, N. Podolefsky, M. Dubson, N. D. Finkelstein and C. E. Wieman, A new instrument for measuring student beliefs about physics and learning physics: the Colorado Learning Attitudes about Science Survey, *Physical Review Special Topics - Physics Education Research* **2**, 010101 (2006).
- [37] L. McCullough, Gender, context, and physics assessment, *Journal of International Women's Studies* **5**, 20 (2004).
- [38] T. L. McCaskey, M. H. Dancy and A. Elby, "Effects in assessment caused by splits between belief and understanding" in *Proceedings of the 2003 Physics Education Research Conference*, edited by J. Marx, S. Franklin and K. Cummings (AIP Conference Proceedings, New York, 2004) p. 37.
- [39] E. Hazel, P. Logan and P. Gallagher, Equitable assessment of students in physics: importance of gender and language background, *International Journal of Science Education* **19**, 381 (1997).
- [40] Note that not all students who enrolled in the course received a course grade. 163 students withdrew from the course (receiving a grade of W) and 2 students took the course pass/fail (receiving a grade of P). Both students who took the course pass/fail took the FMCE. All but one of the students who withdrew from the course did not take the FMCE.
- [41] Updated results, which include data from all semesters between Spring 2004 and Fall 2009, can be found in the Appendix.
- [42] The error bars in all figures represent the standard error on the mean. They are only a lower limit on the actual error. These error bars do not account for sources of error other than statistical error, such as systematic error or sampling bias.
- [43] The fully interactive courses have a higher average learning gain than the partially interactive courses for all students ($\langle g \rangle_{\text{partial}}=0.45$, $\langle g \rangle_{\text{full}}=0.52$).
- [44] A. F. Heckler, Measuring student learning by pre and post testing: absolute gain vs. normalized gain, Unpublished. Can be found at: <http://www.mps.ohio-state.edu/Personnel/Heckler/TauVsGpaper.final.pdf>.
- [45] The CLASS that was administered in semester A1-1 was the old version. Because of the many differences between the old and new versions, we excluded data from that semester from this analysis and future analyses where the CLASS was used.
- [46] Shifts towards less expert-like beliefs are common, as documented in reference 36.
- [47] R. G. Lomax, *An Introduction to Statistical Concepts for Education and Behavioral Sciences* (Lawrence Erlbaum Associates, Mahwah, NJ, 2001).

- [48] A nontrivial fraction of the students score high on the pre-test, making their gain questionable (ceiling effect). The distribution of FMCE post-test scores is not Gaussian, and is skewed towards high scores (non-normal data). The variance in the post-test scores is not constant across all pre-test values (heteroskedasticity). The sample of students used in the multiple regression analysis was not randomly chosen (non-random sampling).
- [49] R. A. Berk, *Regression Analysis: A Constructive Criticism* (SAGE Publications, Thousand Oaks, CA, 2004), p. 212.
- [50] Complete data means that data were available for at least one of the four math tests and all other variables in Table 1 and Table 4. Also, semester A1-1 was excluded from both regression analyses, see Reference 45.
- [51] W. Mendenhall and T. Sincich, *A Second Course in Statistics: Regression Analysis* (Pearson Education, Inc. Upper Saddle River, NJ, 2003).
- [52] The inclusion of the interaction term in Model 3 changes the relationship between the odds for a female and the odds for a male. The relationship becomes $odds_F = \exp(-0.534 + 0.013 \times FMCEPRE) \times odds_M$.
- [53] An R^2 value of 0.45 is relatively high for regression analyses such as this. Similar analyses conducted by other researchers have R^2 values between 0.16 (Tai and Sadler) and 0.36 (Sadler and Tai).
- [54] S. J. Spencer, C. M. Steele and D. M. Quinn, Stereotype threat and women's math performance, *Journal of Experimental Social Psychology* **34**, 4 (1999).
- [55] M. Dancy, "The myth of gender neutrality" in *Proceedings of the 2003 Physics Education Research Conference*, edited by J. Marx, S. Franklin, and K. Cummings (AIP Conference Proceedings, New York, 2004) p. 31.
- [56] B. D. Tatum, *Why are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race* (Basic Books, New York, 1997).
- [57] V. Valian, *Why So Slow: The Advancement of Women* (The MIT Press, Cambridge, 1999).

CHAPTER 3: GENDER DISPARITIES IN PHYSICS 2

I. Introduction

In the current study, we continue our work in characterizing gender differences in participation and performance by examining the second-semester, calculus-based electricity and magnetism (E&M) course. This course is particularly interesting as fewer students (male or female) have significant exposure to the E&M content before coming to the course. As discussed more below, there are several differences between the introductory mechanics course and the introductory E&M course. With the present study of gender differences in the E&M course, we begin to tease apart some of the similarities and differences between performance in Physics 1 and Physics 2, and to get a sense of which, if any, contextual factors differentially impact males and females. In this chapter, we address the following research questions: 1) What fraction of students, and of physics majors specifically, are retained from Physics 1 to Physics 2? Are there differences between students who continue in the introductory sequence and those who do not? 2) How do the performance, attitudes and beliefs, and preparation of males and females in the second-semester introductory physics course compare? and 3) To what extent can prior factors help explain or account for the persistence of the gender gap in the second-semester class?

We begin by briefly reviewing the relevant prior literature and background information (see Chapter 2 for a more thorough review). We then outline the research methods that were used for this study including a description of the Physics 2 course, the student population in this course, the measures of performance and student background that were examined and the sample of students used in this study. We then present the results in four sections, roughly corresponding

to the three major research questions. In summary, we find no differences in the retention rates of males and females from Physics 1 to Physics 2 for students overall, though there are small differences in retention rates of male and female physics majors, with males systematically more likely to continue and less likely to drop out than females. The trends we see in terms of male and female course grades in Physics 2 match those observed in Physics 1 except for some notable cases in which males and females have significantly different course grades. Despite males and females having similar E&M conceptual pre-test scores at the beginning of Physics 2, males outperform females at the end of the semester by about 6 percentage points. This post-test gender gap can largely be attributed to differences in males' and females' prior physics performance (FMCE post-test, BEMA pre-test, and Physics 1 exam grades), mathematics standardized test performance, and students' attitudes and beliefs. A multiple regression model of students' conceptual performance suggests these prior factors can account for up to 60% of the observed gender differences. Taken all together, our current study of the second-semester physics course indicates that there is not one single factor that can explain the under-representation of females in physics, but it is rather the building up of small differences between males and females over time that may be responsible for the large disparities in participation of males and females in physics.

II. Background

In their study of the relationship between post-course gender gaps and the level of interactive engagement in a course, Lorenzo, *et al.* [1] argued that the more interactive a class is, the smaller the post-course gender difference will be. They also argued that the reduction of the gender gap is independent of the instructor, although we found that the instructor and the choices that the instructor makes may impact the gender gap (see Chapter 2). In addition to the use of

interactive techniques and the instructor, there are other contextual factors in the classroom that can contribute to the gender differences we observe in the course, such as content covered, student demographics, climate, or how different activities are framed. At this point, it is unclear which of these factors are critical in reducing or eliminating gender disparities. To further explore some of the contextual factors that may be key, in this study we examine in detail the gender differences in the second-semester, E&M course.

The second-semester introductory physics course is different in many ways from the first-semester course. Most apparent are the differences in physics content. The first-semester course (Physics 1) covers mechanics, including Newton's laws, work, energy, momentum, and waves. The second-semester course (Physics 2) covers electricity and magnetism, including electric fields, Gauss's law, circuits, magnetic fields, and EM-waves. Not only is the content covered in the two courses different, but student familiarity with the content also varies between the two courses. As reported in Chapter 2, 72% of students in Physics 1 have taken one year of high school physics, most likely a mechanics course. Only 14% of students have taken two years of high school physics, suggesting that only a small fraction of students may have seen much E&M content in high school. Another way in which Physics 2 differs from Physics 1 is in the student population, which will be addressed in more detail later. Notably, there are fewer non-science and undeclared majors in Physics 2 than in Physics 1. Though *Tutorials* are used in both Physics 1 and 2, there are a greater number of individual tutorials in Physics 2 that require the use of equipment. Several studies have found differences in how male and female students engage with lab equipment [2,3,4], and these differences could have more of an impact on student learning in Physics 2 than in Physics 1.

While there has been a lot of interest in looking at the performance of males and females in physics, most of the research studies [1,5,6,7,8] have focused on the first-semester, mechanics course. One exception is Meltzer's study of a "hidden variable" in electricity and magnetism conceptual test performance [9]. Meltzer found that while students' conceptual pre-test scores were not correlated with their normalized learning gains on the Conceptual Survey in Electricity (CSE), students' pre-instruction mathematics skill was correlated with CSE gains, suggesting that differences in learning gains between two populations may be due in part to different incoming math skill, rather than different pre-test scores or abilities to learn physics concepts.

III. Research Methods

A. Course Description

The data in the following studies were collected from ten offerings (between Fall 2004 and Spring 2009) of the second-semester calculus-based introductory electricity and magnetism (E&M) course at the University of Colorado (CU). These are large-enrollment courses that typically have 300 to 500 students. All ten classes used interactive engagement (IE) techniques. Each of the ten classes employed student discussions around ConcepTests (*Peer Instruction*) [10] in lecture, online homework systems [11], and voluntary help-room sessions on problem-solving homework. In addition, all ten classes used *Tutorials in Introductory Physics* [12] and Learning Assistants [13] during a 1 hr/week recitation. There is no laboratory associated with this course. A more detailed description of the course structure can be found in prior work [14]. In the previous chapter on the first-semester course, we categorized courses as IE 1 (partially interactive, no *Tutorials*) or IE 2 (fully interactive, use *Tutorials*). All ten classes in this study are categorized as IE 2. Though we categorize all of these classes as IE 2, we also recognize that there are a variety of faculty teaching these classes who have differing levels of experience and

familiarity with the interactive engagement methods that are employed. Though the curriculum looks the same, we know that how the curriculum is enacted can be very different [15].

The ten classes included in this study were taught by seven different instructors [16]. Each semester included in the study will be identified by a code. All codes will have a form similar to A1-2, where the letter represents the lead instructor, the 1 indicates that this is a semester of Physics 1, and the 2 indicates that it is the second time that Professor A has taught this course in the semesters included in our data set. All instructors except one (Professor N) were male.

B. *Student Population*

The student population in the second-semester introductory course is about one-quarter female, just as is the case in Physics 1. Over half of the students are declared engineering majors and about 20% are other science majors. Only about 8% of the students who are enrolled in the introductory E&M course are declared physics majors. This is a slightly different student population than we see in Physics 1. Not surprisingly, there are fewer non-science and undeclared majors in Physics 2 than in Physics 1. Also, a larger fraction of the students are engineering majors in Physics 2 than in Physics 1. We also see significant differences in student major by gender [$\chi^2(4, N=3894)=140.97, p<0.01$], as shown in Table 1. Namely, females are less likely to be engineering majors, and about twice as likely to be other science majors, which is similar to trends observed in Physics 1. Unlike in Physics 1, where the percentages of male and female physics majors were not different (5.6% of males and 5.2% of females), in Physics 2 a higher fraction of the male students are physics majors as compared to female students (9% of males versus 6% of females). This difference is significant [$\chi^2(2, N=3894)=8.14, p<0.01$]. Looking at student ethnicity, over 80% of the students are white, about 9% are Asian, and less

than 10% are African American, Hispanic, or Native American. There are only small differences in ethnicity by gender. Further, there are no differences between the ethnicity distributions of students in Physics 2 compared to Physics 1.

Table 1. Frequencies for gender, student declared major, and ethnicity for all students in the study, that is, students who enrolled and received a grade in the second-semester introductory physics course between Fall 2004 and Spring 2009.

Gender (N=3895)	%		
Male	75.3		
Female	24.7		
Major (N=3894)	%	% of males	% of females
Physics	8.2	9.0	6.0
Engineering	57.6	60.3	49.1
Other Science	21.8	17.6	34.7
Non-Science	5.3	5.0	6.4
Undeclared or Other	7.0	8.1	3.9
Ethnicity (N=3662)	%	% of males	% of females
Asian	8.7	7.8	11.6
African American	1.3	1.2	1.4
Hispanic	6.0	6.0	6.0
Native American	0.7	0.7	0.7
White	81.4	82.6	77.7
Foreign	2.0	1.8	2.5

C. Performance and Background Measures

Of primary interest in this study is to what degree males and females differ on measures of background and preparation and to what degree these differences contribute to the observed gender gap. Conceptual performance, as measured by the Brief Electricity and Magnetism Assessment (BEMA) [17], serves as the focus of the study. The BEMA post-test score for each student is used as a measure of the student's conceptual understanding of physics at the end of the semester. Only students with matched pre-test and post-test data are included. In two of the ten semesters the BEMA was not given to students as a pre-test, and was only given at the end of the semester as a post-test [18]. Though we have post-test data for these semesters, we do not

include them in most of the analyses in this chapter, as we cannot match individual students' pre- and post-test scores. In another two of the ten semesters, the BEMA was only given to half of the students in the class. We also have excluded these semesters in some of the analyses in this chapter. The number of students with matched pre- and post-test data from the remaining six semesters is 1704. Additional evaluation of student performance in the course is captured by homework, exam, participation, and course grades, which were collected from the instructor in each semester.

Data have been gathered [19] on students' background knowledge and their preparation for college physics. Prior academic performance is captured by students' high school grade point average (GPA), while the BEMA pre-test scores and FMCE post-test scores from the previous term (when available) are used to measure students' prior conceptual understanding of physics. The math portion of the Scholastic Aptitude Test (SAT-math) and the math portion of the American College Test (ACT-math) were used as measures of students' prior knowledge of mathematics [20]. Scores on each of the math tests were similarly correlated with the BEMA post-test ($r \approx 0.35$) and were also high correlated ($r = 0.71$) with each other (for the almost 2000 students who took both tests). To get a measure of prior math knowledge for almost every student and to avoid having multiple variables that contained the same information, the scores on the two tests were combined. The scores for each test were first normalized (converted to z -scores [21]). For students who took only one of the two tests, the z -score on that test was used to measure mathematics knowledge. For the smaller number of students who took both tests, the combined math score is an average of the z -scores for each test. Student course preparation for college physics is measured by how many years of high school physics and calculus a student

had taken. Data were not available on the grade that students received in their high school courses.

Just as in the prior chapter, data were also collected on students' attitudes and beliefs about physics and about learning physics. Attitudes and beliefs are measured by the Colorado Learning Attitudes about Science Survey (CLASS) [22]. Favorable pre-test scores on each category are used as measures of students' incoming beliefs. Favorable post-test scores and shifts (*post – pre*) are used as measures of students' attitudes and beliefs at the end of the semester and to measure change in attitudes and beliefs, respectively. The CLASS is administered at the beginning and end of Physics 1 and Physics 2. For those students who took both the pre- and post-CLASS in Physics 1 and Physics 2, the correlation between the post-Physics 1 CLASS score and the pre-Physics 2 CLASS score is 0.73, suggesting that students' attitudes and beliefs are fairly stable over the winter and summer breaks. This analysis will primarily use results from the survey administered in Physics 2.

D. Study Sample

The BEMA is administered the first and last weeks of the semester during recitation, and only those students that attend both weeks take the pre- and post-BEMA. This non-random sampling could introduce bias into our results. To understand the bias of our sample we compare students who did and did not take both the pre- and post-BEMA. Of the 2318 students who took Physics 2 during the semesters included in this study (and in the semesters where the BEMA was offered to all students both pre- and post-instruction), 1704 students (74%) took both the pre- and post-BEMA. Comparing the populations who did and did not take both the pre- and post-BEMA, we find that females were more likely to take the BEMA than males: 80% of females took the BEMA, while only 72% of males took the BEMA. The course grades (on a scale from 0.0 to 4.0)

for males and females in each group are shown in Table 2. The average course grades of students who took the BEMA are higher than the course grades of students who did not take the BEMA [males: $t(1755)=18.66$, $p<0.01$; females: $t(559)=8.90$, $p<0.01$]. While this is a source of bias, there is no significant gender gap in course grades for either of the two groups [with BEMA: $t(1702)=1.02$, $p=0.31$; without BEMA: $t(612)=-0.29$, $p=0.77$]. By focusing on the BEMA as a measure of learning, we limit the sample of students included in the analysis and exclude primarily those with lower course grades. But, the similarity in gender gaps in course grades for the two groups suggests that the estimate of the gender differences in conceptual performance provided by the BEMA may be a reasonable estimate of the gender difference for all students.

Table 2. Average course grades for males and females who did and did not take the BEMA. Course grades are on a 0.0 – 4.0 scale. The * indicates that the difference is statistically significant at the $p<0.01$ level.

	Males			Females			M – F
	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	
Students without pre- and/or post-BEMA	500	1.93	1.1	114	1.96	1.2	-0.03
Students with pre- and post-BEMA	1257	2.85	0.8	447	2.80	0.8	0.05
BEMA – no BEMA		0.92*			0.84*		

IV. Results

A. Tracking Students from Physics 1 to Physics 2

Before looking at differences between males and females in the second-semester course, we look at which students continue from Physics 1 to Physics 2 and whether there are differences by gender. For this analysis, we only included students who took Physics 1 between Spring 2004 and Spring 2008. We find that the majority of students who do take Physics 2, take it within a year of taking Physics 1. By only looking through Spring 2008, we ensure that the majority of students included in our analysis will have taken Physics 2 if they were likely to do so. Figure 1

shows the number of males and females (physics majors are listed in parentheses) at each step of the progression from Physics 1 to Physics 2. Of the students who took Physics 1, 37% of both males and females did *not* take Physics 2. We see no gender difference in the percentage of students who do not go on to Physics 2. Of the students who took Physics 2, about 20% of both the males and females did not take Physics 1. Looking both at the number of students who drop out of the introductory sequence and who join in the sequence after Physics 1, we find no gender differences.

Next, we concentrate on physics majors, that is, those students who were declared physics majors the semester that they took Physics 1 or 2, regardless of whether they remained physics majors until graduation. As we mentioned above, there is not a significant difference in the percentage of male and female physics majors who take Physics 1. Also, there is not a significant difference ($p>0.8$) in the percentage of males and females who were declared physics majors in Physics 1, but who never took Physics 2, and presumably changed their major. 25% of female physics majors and 23% of male physics majors in Physics 1 never took Physics 2. On the left side of Figure 1 is information about students who took Physics 2, but who changed their major between Physics 1 and Physics 2. Of students that were declared physics majors in Physics 1, 11% of the females and 7% of the males changed their major to something other than physics between Physics 1 and 2 (not a significant gender difference). Of students that were not declared physics majors in Physics 1, 0.4% of females and 0.8% of males switched their major to physics between Physics 1 and Physics 2 (not a significant gender difference). These small but consistent gender differences in the numbers of students who continue through the introductory sequence result in a significantly smaller percentage of females who are physics majors in Physics 2 compared to males. In Physics 2, 6% of the females are physics majors and 8% of the males are

physics majors. Looking at the trajectories of males and females from Physics 1 to Physics 2, we see no significant gender differences, but the small, non-statistically significant differences combine such that there is a smaller percentage of females than males who are physics majors in Physics 2.

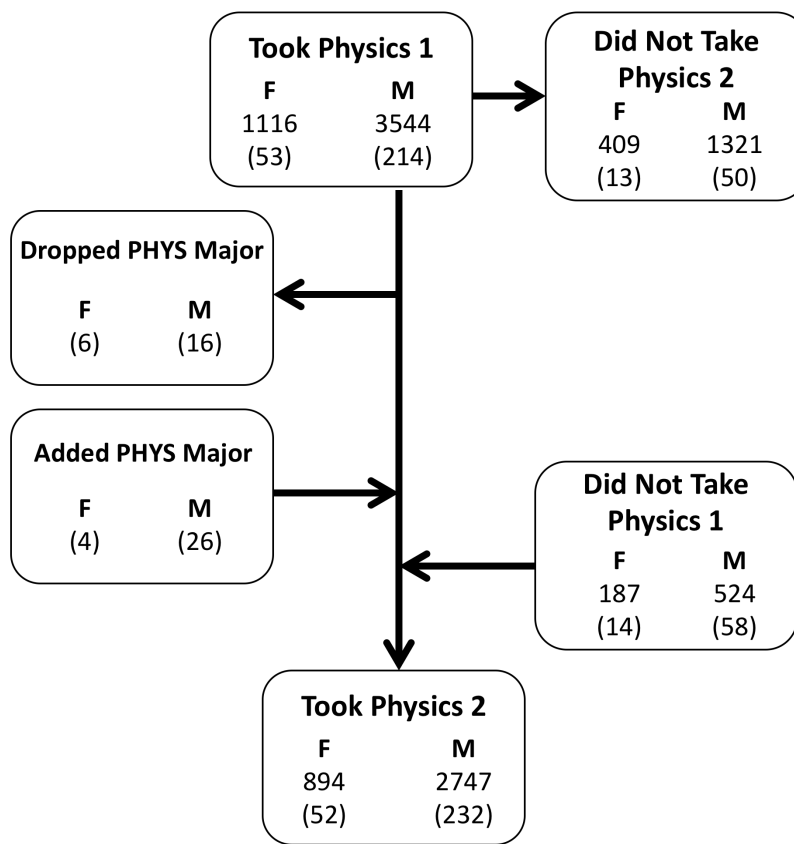


Figure 1. Tracking students through the introductory physics sequence. The chart above shows the numbers of males and females who took Physics 1 (between spring 2004 and spring 2008) and Physics 2 (between fall 2004 and spring 2009). The numbers in parentheses are the number of male and female *physics majors* (PHYS) at each step. Males and females are about fractionally equal at every step of the chart, except in the percentage of physics majors in Physics 2.

We also compare the Physics 1 course grade and FMCE scores of students who did and did not go on to Physics 2. These comparisons are found in Table 3. We begin by comparing students within each gender. Not surprisingly, the students who did take Physics 2 have higher course grades in Physics 1 than the students who did not take Physics 2. The difference is

significant [males: $t(3542)=23.77, p<0.01$; females: $t(1114)=14.05, p<0.01$] and about the same for both males and females. The effect sizes [23] of the differences are 0.77 for males and 0.81 for females, both relatively large effect sizes. Having found sizeable differences between the course grades of students who did and did not take Physics 2, we now look at FMCE scores, beginning with the pre-test. Males who did take Physics 2 had significantly higher [$t(2110)=6.35, p<0.01$] FMCE pre-test scores than males who did not take Physics 2. The difference was about 6% (an effect size of 0.29). However, the FMCE pre-test scores of females who did and did not take Physics 2 are not significantly different [$t(733)=1.67, p=0.10$]. This suggests that despite the two groups of females being equally (un)prepared for Physics 1 in terms of incoming conceptual performance, some of the females continued on in physics while other females did not. It appears that FMCE pre-test score is an indicator of whether males move on to Physics 2, but it is not an indicator for females. There are also differences on the FMCE post-test score for both males and females comparing those who did and did not take Physics 2 [males: $t(2110)=8.23, p<0.01$; females: $t(733)=3.96, p<0.01$]. The post-test differences between those students who only took Physics 1 and those students who went on to Physics 2 are larger for both males (11%) and females (9%) than were the pre-test differences. The effect sizes of the post-test differences are 0.38 for males and 0.31 for females.

We now look at the gender differences for students who do and do not go on to Physics 2. We might expect that for those students who continued on in the introductory physics sequence, we would not observe the same gender differences that we observed for the Physics 1 class as a whole, that is, the gender differences observed in Physics 1 may be primarily due to those students who drop out of the introductory sequence by Physics 2. We find that this is not the case in general. In terms of course grade, the gender difference in course grade for students who did

take Physics 2 is slightly smaller than the gender difference for the students who did not take Physics 2. But, the situation is reversed when looking at the FMCE pre- and post-test. The gender gap on FMCE pre- and post-test for students who did take Physics 2 is larger than for students who did not take Physics 2. The gender gap (and its increase from pre- to post-test) persists even when only looking at this special subpopulation of students who continued on to Physics 2.

Table 3. Gender gaps in course grades and FMCE for those students who took Physics 1 but then did and did not take Physics 2. The differences in the bottom row are (Physics 1 and 2 – Physics 1 only). The asterisk (*) indicates that the difference is significant at the $p < 0.05$ level.

	Physics 1 Course Grade			FMCE Pre-test			FMCE Post-test		
	M	F	M – F	M	F	M – F	M	F	M – F
Physics 1 only	2.03	1.90	0.13*	28.1	21.0	7.1*	59.0	48.1	11.1*
Physics 1 and 2	2.83	2.74	0.09*	34.4	23.1	11.3*	69.6	57.0	12.6*
Differences	0.80*	0.84*		6.3*	2.1		10.6*	8.9*	

In addition to looking at performance measures to compare students who do and do not go on to Physics 2, we can also look at students' attitudes and beliefs, as measured by the CLASS instrument. In Table 4 we present the CLASS pre- and post-test scores for males and females who did and did not take Physics 2. These data are collected at the beginning and end of Physics 1. As with prior studies, we observe an overall negative shift in student attitudes and beliefs for all students. Again, we might expect that those students who go on to Physics 2 would have more favorable attitudes and beliefs than those students who do not go on. We do find that students who take Physics 2 have more favorable attitudes and beliefs both at the beginning and end of the semester than students who do not take Physics 2 (though the difference in pre-test scores for females is not significant). Though these differences between students who do and do

not take Physics 2 are significant, the effect sizes of the differences are small, between about 0.1 and 0.3.

Table 4. Gender gaps in Physics 1 CLASS (% favorable) pre-test, post-test, and shifts for those students who took Physics 1 but then did and did not take Physics 2. The differences in the bottom row are (Physics 1 and 2 – Physics 1 only). The asterisk (*) indicates that the difference is significant at the $p < 0.05$ level.

	Physics 1 CLASS Pre-test			Physics 1 CLASS Post-test		
	Males	Females	M – F	Males	Females	M – F
Physics 1 only	62.7	62.0	0.7	56.0	52.3	5.7*
Physics 1 and 2	66.4	64.2	2.2*	61.3	57.5	3.8*
Differences	3.7*	2.2		5.3*	5.2*	

In summary, despite the gender differences that we see at the end of Physics 1 (in terms of FMCE post-test score), we find that males and females are continuing through the introductory sequence (and not continuing) at the same rate. The same is true of the physics majors, though we do see a smaller percentage of female physics majors in Physics 2 compared to males; 6% of females versus 8% of males are declared physics majors in Physics 2. As evidenced by the Physics 1 grades and FMCE post-test scores of males and females who continue on to Physics 2, the females who are taking Physics 2 are less prepared than the males. Females who continue in the introductory sequence also have less favorable attitudes and beliefs than the males who take Physics 2. Having examined the gender differences in retention from Physics 1 to Physics 2, in the next section, we take a closer look at the performance gender differences in the second-semester physics course.

B. *Identifying Differences by Gender*

1. College Course Performance Differences

We now focus our attention on students in the second-semester introductory course by examining conceptual mastery, course grades, DFW rates (grades of D, F or withdrawal), and attitudes and beliefs.

a. Conceptual surveys

We first look at students' conceptual performance as measured by the BEMA. Figure 2 presents the pre- and post-test gender gaps for each semester included in the study. Recall that in two of the ten semesters (Semesters M2-1 and G2-3) students were given the BEMA only at the end of the semester as a post-test. We include the post-test gender gaps for these semesters, even though we have no pre-test data. In five of the eight semesters of pre-test data there is not a significant gender difference in pre-test scores [t 's < 1.95, p 's > 0.05]. Males and females do not score significantly differently on the BEMA pre-test in the majority of the semesters that the BEMA pre-test has been given. In the remaining three semesters, where there is a statistically significant pre-test difference, the gender gap is only between 2.6% and 3.6%. This is much smaller than the gender gaps that we observe on the FMCE pre-test at the beginning of Physics 1, which are between 6% and 14% (about 10% on average). Taking the BEMA pre-test score as a measure of preparation, it seems that males and females are equally prepared for Physics 2 in terms of exposure to E&M content.

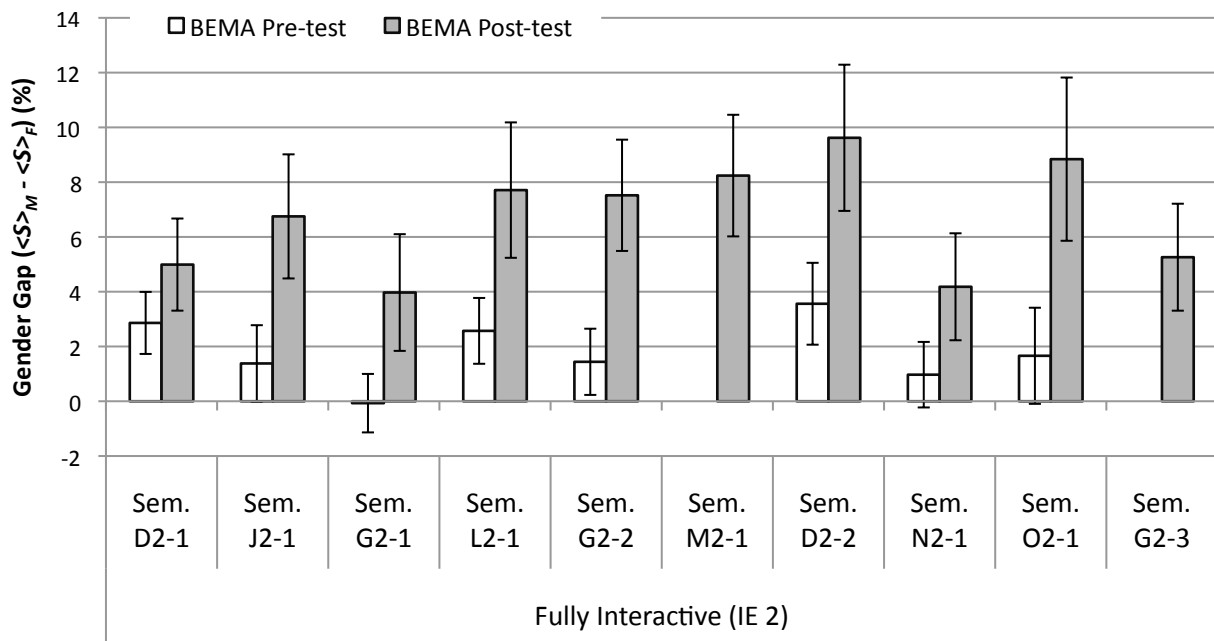


Figure 2. Pre- and post-test gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) by semester. The data shown here include all students who took the pre- and post-BEMA. These data represent seven different instructors and over 2500 students. The semester is indicated along the x-axis. The error bars represent the standard errors of the mean.

Despite equal preparation of males and females, the BEMA post-test gender gap is statistically significant in all nine of the ten semesters [t 's > 2.12, p 's < 0.04]. In semester G2-1, the BEMA post-test gender gap is only marginally significant [$t(316) = 1.89$, $p = 0.06$]. Males scored significantly higher on the BEMA post-test than females in all semesters. Even in those semesters where there was no significant difference on the pre-test, males and females performed differently on the post-test. Recall that all ten semesters used fully interactive engagement methods, including *Peer Instruction* [10] and *Tutorials* [12]. Despite the use of these IE methods, the gender gap increases from pre- to post-test in all semesters of Physics 2. The post-test gender gap ranges from 4.0% to 9.6%. On average, the effect size [24] of the pre-test gender gap is 0.17, and the effect size of the post-test gender gap is 0.39.

The normalized gain [25],

$$\langle g \rangle = \frac{\langle post \rangle - \langle pre \rangle}{100 - \langle pre \rangle},$$

for these classes is between 0.33 and 0.48. Over the eight semesters in which both the pre- and post-test were given, the average normalized gain is 0.40. These gains match the range of normalized learning gains reported for classes at other institutions that use the Matter and Interactions [26] curriculum [27]. While the normalized gains for the course are in line with gains of other reformed courses, we do see differences by gender. The individual normalized gains [28] of males and females are statistically different [$t(1998)=6.99, p<0.01$]. Females have an average normalized gain of 0.35 (over all semesters) while males have an average normalized gain of 0.42. It appears that females learn a smaller percentage of what they did not already know coming into Physics 2 than males.

From Figure 2, it appears that a gender gap is created over the course of Physics 2. Males and females come into the course with the same level of E&M conceptual understanding, and at the end of the course, the males are performing better on the BEMA than the females. However, if we look at the FMCE post-test gender gaps for those students who took both the BEMA and the FMCE, shown in Figure 3 [29], we find that in most semesters the post-FMCE gender gaps are larger than the post-BEMA gender gaps. Further, in our classes the average pre-test scores on the BEMA are 25.3% for females and 26.8% for males. These scores are close to “informed guessing”. Most students only take one year of high school physics, which is most likely a mechanics course, so most students, male or female, probably have not been exposed to much E&M when they come into Physics 2. One interpretation is that the pre-test gender gap is masked by lack of conceptual exposure to the subject. The BEMA does not measure what students know on the pre-test, and in fact, if the FMCE is taken as the measure of conceptual

performance upon entering Physics 2, it appears that we may be reducing the gender gap from Physics 1 to Physics 2.

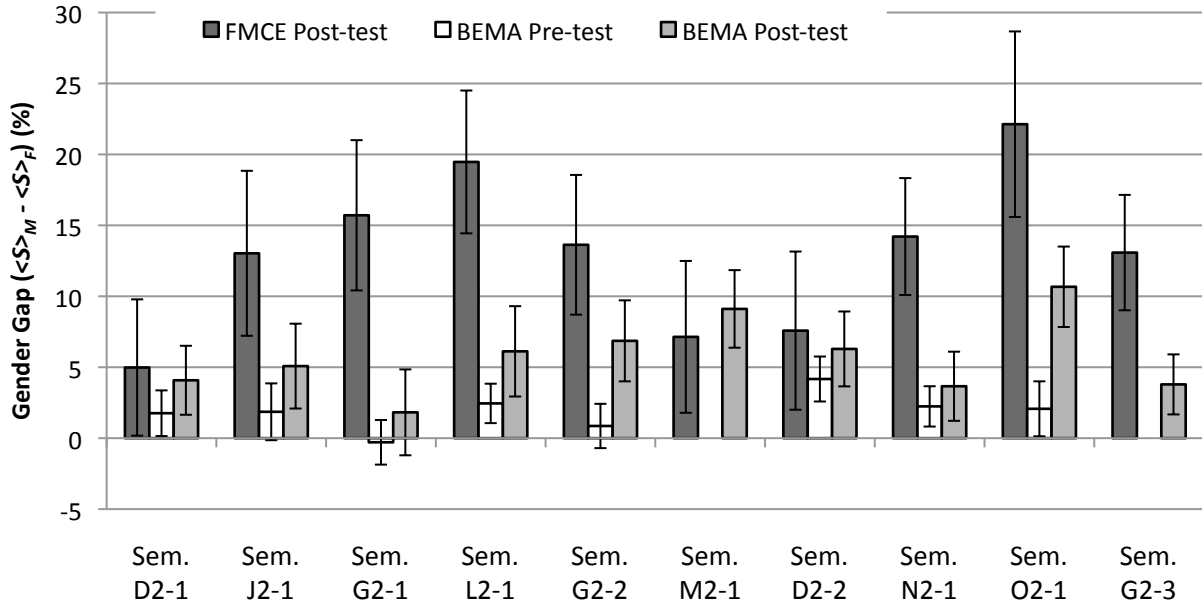


Figure 3. FMCE and BEMA gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) by semester. The gender gaps for each semester are shown above ($N=1487$) for only those students who took both the FMCE (in the first-semester course) and the BEMA (in the second-semester course). The error bars represent the standard errors of the mean.

b. Course grades

In addition to looking at student performance on conceptual surveys, we can also look at how males and females performed in the course overall and on each of the components of the course. For each of the ten semesters of the E&M course males' and females' scores are averaged on participation, homework, exams, and total course grade. In all of these courses, exams make up 60% to 70% of the course grade, homework counts for 25% to 35%, and participation makes up the remainder, between 0% and 10%. The difference between the average scores of males and females in each component ($\langle S \rangle_M - \langle S \rangle_F$) is calculated for each semester. These differences for each semester, along with the average differences across all semesters, are shown in Table 5.

On average in Physics 2, females outscore males by about 6% on participation and by about 5% on homework, but males outscore females by about 4% on exams. This is very similar to what we reported in studies of Physics 1. These differences on participation, homework, and exams offset one another resulting in course grades of males and females that are not statistically different [$t(3893)=-0.05$, $p=0.96$]. This happens on average and in most individual semesters. There are two notable differences from this trend. In Semester O2-1, the differences in males' and females' homework and participation scores are on the smaller side, and the difference in exam scores is the largest that we report. This results in course grades of males and females that are significantly different [$t(427)=2.09$, $p=0.04$]. Males have course grades about 0.2 grade points higher than females (on 0 to 4.0 scale).

Another semester that stands out is Semester N2-1, the only semester in which there was a female professor. In this semester females have much higher participation and homework scores than males. The differences of 12% on participation and 8% on homework are the largest differences in this data set. In addition, the exam scores of males and females are not significantly different. This leads to females having higher course grades than males by about a third of a letter grade. This is the largest gender difference that we have seen in course grades (including Physics 1 and Physics 2), and the only time we have seen females with statistically significantly higher course grades than males [$t(399)=-2.84$, $p<0.01$]. Though females outperformed males in the course overall in this semester, there was still a small (4%), but statistically significant, BEMA post-test gender gap.

Table 5. Analysis of students' course grades. Each column contains the difference between the average scores for males and females ($\langle S \rangle_M - \langle S \rangle_F$). Error (shown in parentheses) is computed from the standard errors of the mean for males and females added in quadrature. The asterisk (*) indicates that the difference is statically significant at the $p < 0.05$ level. In semester L2-1 no participation credit was given.

	Participation (%)	Homework (%)	Exams (%)	Course GPA (4 pt. scale)
Semester D2-1	-6.4 (1.0)*	-7.0 (1.5)*	3.6 (1.6)*	-0.04 (0.12)
Semester J2-1	-6 (3)*	-4 (2)	6.1 (1.9)*	0.08 (0.13)
Semester G2-1	-7.4 (1.6)*	-3.3 (1.8)	4.2 (1.5)*	-0.03 (0.12)
Semester L2-1		-3 (3)	2.5 (1.7)	0.06 (0.13)
Semester G2-2	-7.2 (1.6)*	-5.5 (1.9)*	4.8 (1.7)*	0.07 (0.12)
Semester M2-1	-5 (2)*	-4.0 (1.7)*	2.4 (1.6)	-0.01 (0.13)
Semester D2-2	-4.9 (1.7)*	-2.6 (1.9)	3.1 (1.4)*	0.02 (0.11)
Semester N2-1	-11.5 (1.8)*	-8.2 (1.9)*	-1.5 (1.6)	-0.32 (0.11)*
Semester O2-1	-2.7 (1.9)	-3.7 (1.7)*	7.7 (1.4)*	0.20 (0.09)*
Semester G2-3	-6.1 (1.7)*	-5.3 (1.8)*	4.2 (1.4)*	0.02 (0.11)
Average	-6.1 (0.6)*	-4.7 (0.6)*	3.8 (0.5)*	-0.002 (0.04)

c. DFW rates

Another way to compare course grades is to look at DFW rates of males and females. The DFW rate for each semester is the percentage of students that received a grade of D, F, or W (withdrew from the course [30]). Table 6 lists the DFW rates for males and females in each semester and an average DFW rate over all semesters. We first look to see if there are differences in the DFW rate from semester to semester for males and for females. There are not significant differences in the female DFW rate by semester [$\chi^2(9, N=1030)=10.32, p=0.33$], but there are significant differences in the male DFW rate by semester [$\chi^2(9, N=3160)=27.03, p<0.01$]. We use pair-wise χ^2 tests to determine in which semesters males have significantly different DFW rates. The only significant differences involve semester O2-1 being different from semesters D2-1 and N2-1. Overall, the DFW rates of males and females do not change much from semester to semester, and less than 15% of both males and females receive grades of D, F, or W.

We next compare the DFW rates *between* males and females to determine if they are significantly different in any semester. In most semesters, the DFW rates of males and females are not significantly different [p 's>0.4]. The one exception is Semester N2-1, where the DFW rate for males was 20% and the DFW rate for females was 8%. This is the only semester where the DFW rates of males and females were significantly different [$\chi^2(1, N=431)=8.00, p<0.01$]. We saw above that females had higher course grades in this semester than males. But if we look at just the percentage of males and females who withdrew from the course in Semester N2-1, we find that none of the females withdrew, but 4% of the males withdrew that semester (a significant difference, [$\chi^2(1, N=431)=4.08, p=0.04$]). The only other semester where there was a significant difference in the numbers of males and females that withdrew from the course was semester G2-1, where a higher percentage of females withdrew compared to males [$\chi^2(1, N=444)=4.15, p=0.04$].

Table 6. DFW and W rates for males and females in each semester. The DFW rate is the percentage of students who receive a grade of D, F, or W (withdrew from the course). The W rate is the percentage of students who withdrew from the course. On average, the DFW and W rates of males and females are not significantly different. Standard error on the mean is given in parentheses.

	DFW Rates (%)		W Rates (%)	
	Males	Females	Males	Females
Semester D2-1	17.5	16.7	2.7	4.6
Semester J2-1	16.9	16.7	2.4	5.1
Semester G2-1	15.5	17.6	1.5	4.8
Semester L2-1	12.6	14.7	3.0	2.1
Semester G2-2	12.5	9.3	2.1	4.1
Semester M2-1	15.7	12.5	3.8	2.5
Semester D2-2	16.3	15.2	4.1	2.7
Semester N2-1	20.4	8.4	3.7	0
Semester O2-1	8.5	9.5	3.4	4.3
Semester G2-3	12.3	10.6	0.3	0
Average	14.8 (1.1)	13.1 (1.1)	2.7 (0.4)	3.0 (0.6)

d. Attitudes and beliefs

In addition to looking at performance in the second-semester introductory course, we can also explore how the attitudes and beliefs of males and females change over the course of the semester and whether there are any gender differences. In the previous chapter when looking at the CLASS, we found that both males and females shifted towards less expert-like attitudes and beliefs over the course of the *first-semester* introductory physics course, and females had more negative shifts in all categories than males. This is reflected in the pre-test scores that are collected at the beginning of the second-semester course. In all categories except Sense-Making females have significantly lower average pre-test scores than males ($t(2104) < 2.04$, $p < 0.04$). This means that females come in to the second-semester course reporting less expert-like attitudes and beliefs about physics and learning physics than males. If we look at the shifts in students' attitudes and beliefs over the course of Physics 2, we see different results than we saw in Physics 1. In Physics 1, we found that females had more negative shifts than males in all categories and overall. The shifts that we saw in Physics 1 were between about -5% and -15%. Figure 4 shows the shifts of males and females in Physics 2. We see in Figure 4 that the shifts over the course of Physics 2 are considerably smaller than those from Physics 1, but still zero or negative. The shifts over the course of Physics 2 are between 0% and -6%. We do not find any significant gender differences in the shifts of males and females except in the Personal Interest category [$t(2104) = 2.75$, $p < 0.01$], where females have more negative shifts than males. The difference in shifts in the Sense-Making category is marginally significant [$t(2104) = 1.78$, $p = 0.08$]. The Personal Interest category has one of the largest pre-test gender differences, which, in combination with the gender differences in shifts over the semester, results in an 11% post-test gender difference in the Personal Interest category, larger than any other category. In summary,

the attitudes and beliefs of students do shift towards less expert-like beliefs over the course of Physics 2 [31], but the shifts are much smaller than in Physics 1. Also, there are fewer differences in shifts between males and females in Physics 2 than in Physics 1.

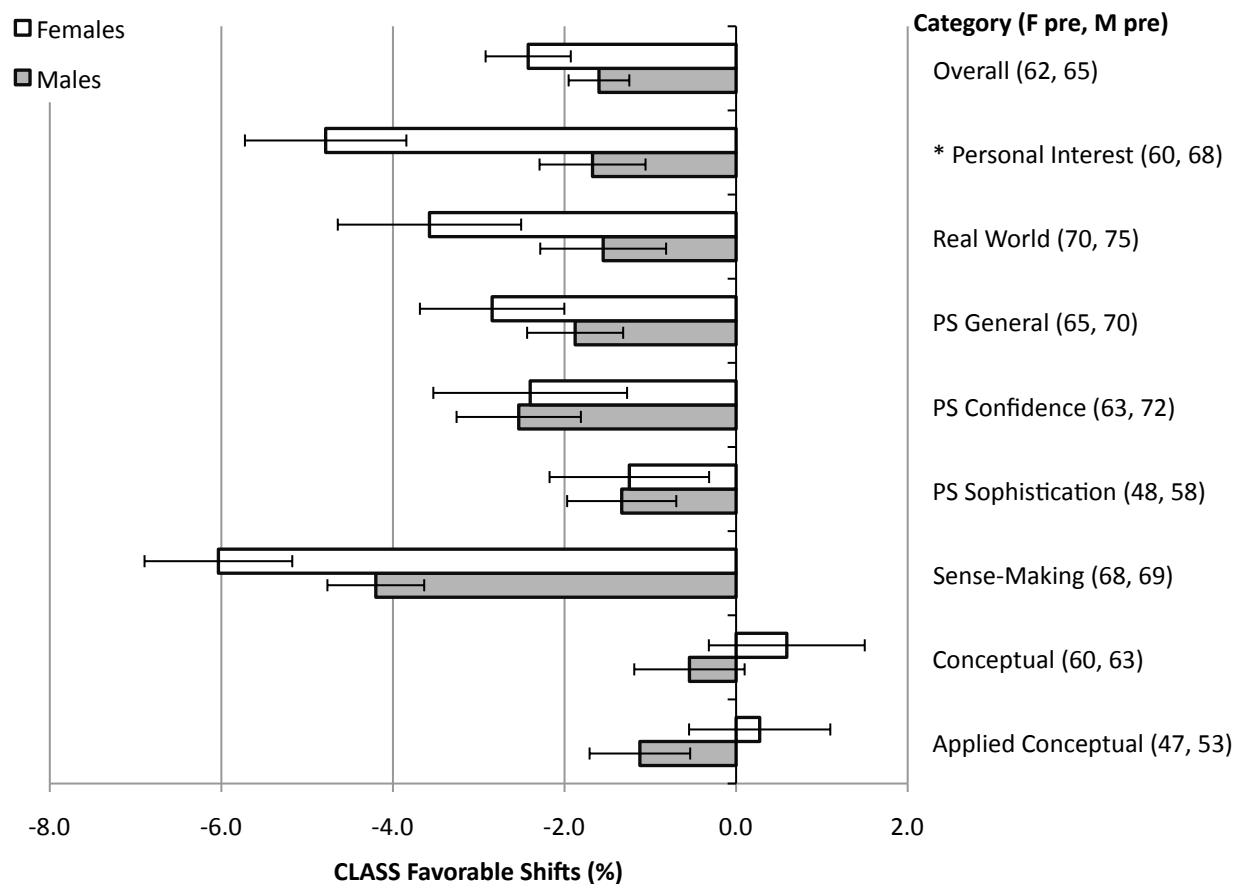


Figure 4. Average shifts (*post – pre*) for males and females in Physics 2 on each of the CLASS categories. All shifts are negative or zero, meaning both male and female students shift toward less expert-like attitudes and beliefs about physics or remain the same. The asterisk (*) indicates that the difference in shifts for males and females is significant ($p < 0.05$). Values in parentheses (on the right hand side) are female and male average pre-test scores. The *pre-test* scores of males and females are significantly different in all categories except Sense-Making.

2. Academic Background Differences

In the previous section we reported the observed differences in males’ and females’ performance and attitudes in the second-semester course. Here, we examine the background and preparation of males and females in Physics 2. As part of students’ background and preparation

we look at both high school factors and data from Physics 1, for those students who took Physics 1. Male and female averages for each of the background variables and the gender differences for each are presented in Table 7 for the population of Physics 2 students. Note that not all data are available for all students, as is the case in any course. As a consequence of missing data the reported averages may be biased due to sampling error. We present them regardless as they are the best estimates we have of the values for all students who enroll in Physics 2.

Just as we saw when comparing the FMCE scores of students who went on to Physics 2 to students who did not in Section IV-A, we find that the measures of students' physics and math background found in Table 7 are higher for this population of Physics 2 students than they were for the population of Physics 1 students (see Table 4 in Chapter 2). Despite the higher level of preparation of these Physics 2 students, Table 7 shows that males have significantly higher [p 's<0.05] values than females on almost all variables. Females take less high school physics than males and score lower on the SAT- and ACT-Math tests. We also see that females perform worse in Physics 1 (as discussed above) than males. Females in Physics 2 had lower FMCE pre- and post-test scores and had lower grades in Physics 1 than males.

Table 7. Male and female average values for variables that were collected. The range of possible scores for each variable is shown in parentheses. The effect size is calculated as $ES = (\langle S \rangle_M - \langle S \rangle_F) / SD$, where the SD for all students is used. Significant differences exist between males and females on all of the variables.

	Males			Females			Differences		
	N	Mean	SD	N	Mean	SD	M - F	ES	p value
High school GPA (0.0 – 4.0)	2816	3.604	0.4	918	3.796	0.3	-0.192	-0.50	<0.01
Years high school physics	2564	1.105	0.6	824	1.021	0.6	0.084	0.15	<0.01
Years high school calculus	2538	0.876	0.6	818	0.899	0.6	-0.023	-0.04	0.33
SAT – Math (200 – 800)	1964	655.2	66	680	640.1	73	15.1	0.22	<0.01
ACT – Math (1 – 36)	2259	28.63	4	791	28.26	4	0.37	0.10	0.01
Math combined (z -score)	2817	0.116	0.9	923	-0.04	1	0.156	0.17	<0.01
FMCE Pre-test (0 – 100)	1575	34.5	22	559	23.2	15	11.3	0.54	<0.01
FMCE Post-test (0 – 100)	1575	70.3	27	559	57.3	28	13.0	0.47	<0.01
Physics 1 Grade (0.0 – 4.0)	2624	2.81	0.8	847	2.71	0.9	0.10	0.12	<0.01
Physics 2 CLASS Pre-test (0 – 100)	1480	65.0	18	625	61.9	18	3.1	0.17	<0.01
Physics 2 CLASS Post-test (0 – 100)	1480	63.4	19	625	59.5	19	3.9	0.20	<0.01
BEMA Pre-test (0 – 100)	1485	26.8	10	515	25.3	9	1.5	0.17	<0.01
BEMA Post-test (0 – 100)	1485	57.7	16	515	51.4	15	6.3	0.39	<0.01

The only background variables in which males do not outperform females are high school GPA, where females outscore males [$t(3732)=-13.57, p<0.01$], and years of high school calculus, where males and females are not significantly different [$t(3354)=-0.98, p=0.33$]. Similar results have been found by other researchers [32,33], who concluded that males and females were equally prepared for the introductory physics courses. We suspect that overall measures of high school grades and enrollment in a calculus course are not as important to performance in an introductory physics course as are enrollment in high school physics (exposure to relevant content) and performance on standardized math tests (measures of mathematics performance). By such metrics, females are less prepared for Physics 1 and Physics 2 than males.

C. Correlation of Student Background with Student Conceptual Performance

Having identified several background variables that vary by gender (high school classes taken, standardized test scores, Physics 1 performance, and BEMA pre-test), we next want to know which of these variables is associated with performance on the BEMA post-test and could potentially account for some of the post-BEMA gender difference that we observe. One way to determine whether a background variable can help account for the BEMA post-test gender gap is to group students according to the background variable and then compare the average BEMA post-test scores of students in each group. In this way, we can control for students background score, only comparing students that are similar on that measure. We would normally begin by looking at the BEMA pre-test, but the lack of spread in BEMA pre-test scores precludes an analysis of this sort. Any reasonable sort yields the same 6% spread that exists in the overall score.

Rather than looking at the BEMA pre-test, we can use the FMCE post-test as a measure of prior conceptual understanding. In Figure 5, we have divided students into five groups by

FMCE post-test score. The groups are divided such that an equal number of students is in each bin ($N \sim 225$). We then calculate the average BEMA post-test score for the males and females in each bin. As is seen in Figure 5, males and females with the same FMCE post-test score have BEMA post-test scores that are not significantly different in all five of the bins [t 's < 1.90 , p 's > 0.06]. Males and females who score similarly on the FMCE post-test in Physics 1, score similarly on the BEMA post-test in Physics 2. Though the difference is not significant in any individual bin, males outperform females in four of the five quintiles. The percentages located above each bar in Figure 5 indicate the percentage of males (or females) that fall into that bin. The distributions of males' and females' among the five bins are not equal. More than half of the females are in the lowest two bins, while just about half of the males are in the highest two bins. The results presented in Figure 5 suggest that by taking into account the FMCE post-test scores of males and females, we can account for a large part of the gender gap in BEMA post-test scores.

A similar analysis could be repeated for each of the background variables in Table 7 separately, but ultimately, we want to know how much of the BEMA post-test gender gap can be accounted for by *all* of the background variables together. We explore this question in the following section.

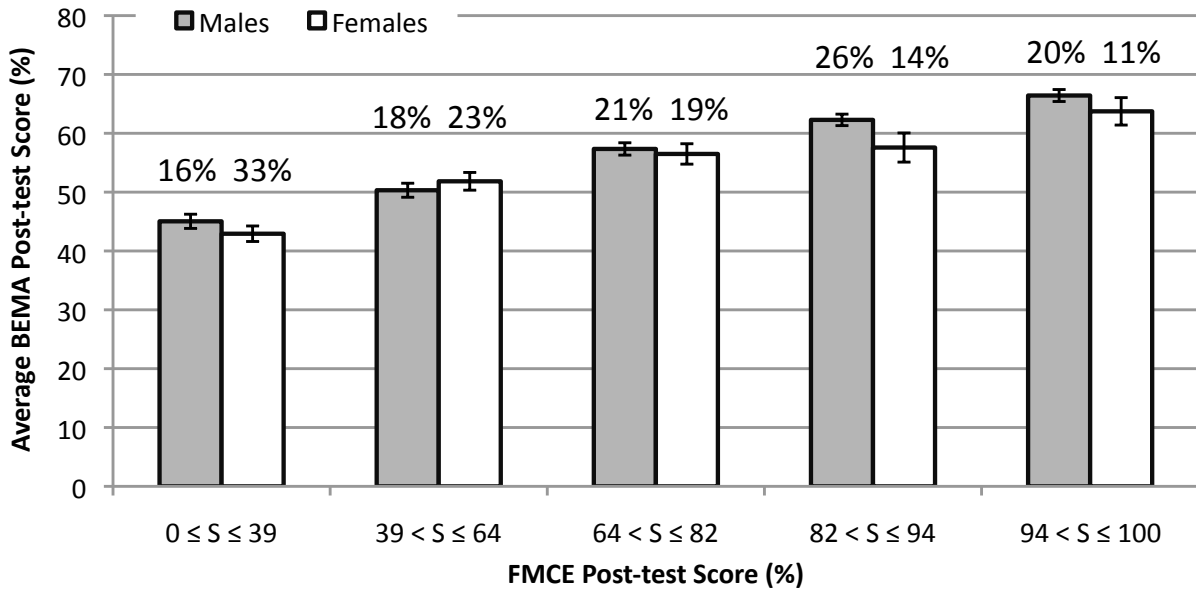


Figure 5. Average BEMA post-test scores for females and males with matched FMCE post-test scores ($N=1117$). The percentages above each bar represent the percentage of the males (or females) from the total in each bin. The error bars represent the standard error on the mean. The differences between males and females are not significant ($p>0.05$) in all five of the bins.

D. Estimation of the Impact of Student Background on the Gender Gap

We investigate whether the background differences between males and females (discussed in the Section IV-B-2) can account for the gender difference that we observe in BEMA post-test scores. We model students' BEMA post-test scores using a multiple regression analysis, which describes the relationship between a student's post-test score and the values of several background variables for that student. Using this relationship, we estimate the difference in post-test scores for males and females with all background variables being held equal. In this way, we will determine how much of the gender gap can be accounted for by factors other than gender.

The post-test scores are modeled according to the equation,

$$BEMAPOST = b_0 + b_1 \times FEMALE + \sum_{k=2}^N b_k \times VAR_k$$

where $BEMAPOST$ is the post-test score on the BEMA, $FEMALE$ is a dummy variable that is 1 for females and 0 for males, and VAR_k are the other background variables that are included in the model and any cross terms between $FEMALE$ and other background variables. b_k are the coefficients for each term, and the multiple regression analysis gives estimates for these coefficients. The coefficient of the $FEMALE$ variable (b_1) gives the difference between a male's and a female's scores, with all other factors being equal. It is this coefficient that we are ultimately interested in.

As in our previous work, we are modeling students' BEMA post-test scores rather than their absolute or normalized gain because we are primarily interested in reducing the gender gap in post-test scores. By modeling the post-test, we can determine what factors influence the post-test score and could therefore contribute to the gender gap. Each of the possible confounding variables is included in the regression analysis. Variables are entered sequentially in order to find the parsimonious combination of factors that best predicts the post-test score for each student. The best model will be judged based on the size of the coefficients, the increase in multiple R^2 (the fraction of variation in post-test scores that is accounted for by the variables in the model), and the significance of variable coefficients.

As stated above, not all data were available for all students. With this being the case, only a subsample of students who took the second-semester introductory course was used in the multiple regression analysis. Recall that only 1704 of the 3895 students who enrolled in Physics 2 during the semesters included in our study took the BEMA pre-test and post-test. Of these 1704 students, complete data (meaning all background variables presented in Table 7) were available for only 637 students. These 637 students make up the first sample used for the analysis. This sample of students is labeled S_I . All of the students in S_I took the FMCE in

Physics 1, so we can use their FMCE post-test score as a measure of their prior understanding of mechanics content. Students' grade in Physics 1 could also be used as a measure of mechanics understanding. If we use Physics 1 grade rather than FMCE post-test score, then we have a second sample of 907 students. This second sample, S_2 , has more students since not everyone takes the FMCE in Physics 1, but everyone receives a grade in Physics 1. We run the regression analysis using both of these samples.

For both samples, it is important to keep in mind that the samples used are not representative of all students who enroll in Physics 2. We can see from Table 8 and Table 9 that the students included in each of the samples have higher course grades than students not in the samples. In all cases, the differences are about half of a letter grade. Though we are sampling students with higher course grades, the gender difference in course grades for both samples is not significantly different from zero, as was the case when looking at the class overall. It appears that the samples used in the regression analyses may be good estimates of the gender *differences* for all students.

Table 8. Average course grades for students included in the first regression sample (S_1) and those who are not in the S_1 sample. Course grades are on a 0.0 to 4.0 scale. The asterisk (*) indicates that the differences are significant ($p < 0.05$).

	Males	Females	M – F
In S_1 sample	2.96	2.92	0.04
Not in S_1 sample	2.46	2.47	0.01
In S_1 – Not in S_1	0.50*	0.45*	

Table 9. Average course grades for students included in the second regression sample (S_2) and those who are not in the S_2 sample. Course grades are on a 0.0 to 4.0 scale. The asterisk (*) indicates that the differences are significant ($p < 0.05$).

	Males	Females	M – F
In S_2 sample	2.93	2.86	0.07
Not in S_2 sample	2.39	2.43	-0.04
In S_2 – Not in S_2	0.54*	0.43*	

If we look further at the BEMA pre- and post-test gender gaps for all students who took the BEMA, students in the S_1 sample and students in the S_2 sample, we see that the gender differences across all three samples of students are very similar. These data are presented in Table 10. This suggests again that the samples used in the regression analyses are reasonable representative of the gender *differences* in the entire population of students.

Table 10. BEMA pre- and post-test gender gaps for all students, students in the first regression sample (S_1), and students in the second regression sample (S_2). The asterisk (*) indicates that the differences are significant ($p < 0.05$).

	All Students			S_1 Sample			S_2 Sample		
	M	F	M – F	M	F	M – F	M	F	M – F
BEMA Pre-test	26.8	25.3	1.5*	26.8	25.0	1.8*	26.7	25.0	1.7*
BEMA Post-test	57.7	51.4	6.3*	58.4	51.7	6.7*	57.5	50.8	6.7*

1. Using Sample S_1

The results of the regression analysis for sample S_1 are shown in Table 11. Three models are reported, starting with a bivariate model that includes only gender and then additional variables are added in each successive model. The table contains the coefficient estimates (b_k) and p values for the coefficients in each model as well as the model-level statistics. The variables that are entered in each successive model are not only significant, but they also increase R^2 substantially (the additional variance explained by each model is significant via F test at the

$p < 0.01$ level). The R^2 for the final model is 0.34, such that the variation in the independent variables explains 34% of the variation in post-test scores [34].

Table 11. Coefficient estimates and multiple regression model statistics for each multiple-regression model. The S_7 sample was used for this regression analysis.

	Model 1		Model 2		Model 3	
Model-level statistics						
Multiple R^2	0.04		0.32		0.34	
F statistic p value	<0.001		<0.001		<0.001	
Residual standard error	15.5		13.1		13.0	
Predictors	b_k	p value	b_k	p value	b_k	p value
Intercept	58.4	<0.01	29.1	<0.01	30.5	<0.01
<i>FEMALE</i>	-6.8	<0.01	-2.6	0.03	-2.6	0.02
BEMA Pre-test			0.25	<0.01	0.27	<0.01
FMCE Post-test			0.24	<0.01	0.23	<0.01
Combined math score			2.1	<0.01	2.0	<0.01
CLASS pre-test			0.07	0.03	0.07	0.04
Semester J2-1					0.78	0.66
Semester G2-1					-4.7	<0.01
Semester L2-1					-2.1	0.27
Semester G2-2					-0.52	0.76
Semester N2-1					-0.08	0.97

We are interested in the difference between males' and females' post-test scores after controlling for several prior factors. In Model 1, where only *FEMALE* is included as an independent variable, the gender difference is 6.8 points. This is just the average difference in post-test scores between males and females in this sample. In Model 2, several covariates that are correlated with the post-test are added. When previous physics performance (BEMA pre-test and FMCE post-test), previous math performance (combined math score), and previous attitudes and beliefs (CLASS pre-test) are controlled for, the gender difference drops to 2.6 points. Already, there is a substantial reduction in the gender difference once previous physics and math performance and attitudes and beliefs are accounted for.

To get a final estimate of the gender difference, we turn to Model 3. In this model, variables are added to take into account the semester that students took Physics 2. The base case in Model 3 is semester D2-1 (meaning there is no variable included for this semester). This means that the coefficients of each semester variable give the average difference between semester D2-1 and that semester after all other variables have been accounted for. For example, controlling for prior physics performance, math performance, and attitudes and beliefs, the average difference between semester D2-1 and semester G2-1 is -4.7 points. This is the only difference that is significant, but this analysis only allows a statistical comparison between semester D2-1 and all other semesters. It does not allow us to compare semesters J2-1 and G2-1, for instance. There could be other significant differences between post-test scores by semester.

With Model 3, a final estimate of the difference between a male's and a female's post-test scores, controlling for several other factors, can be estimated. This difference is 2.6 points. This is a substantial reduction from the 6.8 point difference that is observed just by subtracting the average male and female post-test scores. Controlling for student background in this way, we can account for 62% of the observed gender gap using this final model.

We can also include Physics 1 course grade in the final model, in addition to the FMCE post-test. Though there is an increase in R^2 when Physics 1 course grade is added, because there is not a large gender difference in Physics 1 grade, including it in the model does not lower the coefficient of *FEMALE*, but rather increases it slightly to 3.1 points. We do not include Physics 1 grade in the final model because when it is included, math score and CLASS pre-test are no longer significant predictors of BEMA post-test. Because each of these variables, math performance and prior attitudes and beliefs, are somewhat more explanatory and straightforward

than Physics 1 grade (which is a combination of exams, homework, and participation), we chose to keep them in the final model.

We also attempted to include years of high school physics, students' declared major in Physics 2, ethnicity, and interaction variables between *FEMALE* and all other variables. None of these variables significantly contributed to the model beyond those variables already included in the final model. We suspect that this is primarily due to correlations between these variables and variables already included in the final model.

2. Using Sample S_2

Table 12 presents the results of the regression analysis using the S_2 sample. Recall that for this sample of students we used Physics 1 grade, rather than FMCE post-test, as a measure of prior mechanics conceptual understanding. For this sample, we report four models, starting with a bivariate model and then adding variables in each successive model. The R^2 for the final model is 0.40, such that the variation in the independent variables accounts for 40% of the variation in post-test scores.

Again, we are interested in the coefficient of the *FEMALE* variable. In Model 1, where only *FEMALE* is included, the gender difference is 6.8 points, as we saw above. In Model 2, when covariates are included in the analysis, the gender difference drops to 4.6 points. We note here, that when Physics 1 grade is included in the model, rather than FMCE post-test, less of the gender gap can be accounted for. This is not surprising, since there is not as large of a gender difference on Physics 1 grade as there is on FMCE post-test.

Table 12. Coefficient estimates and multiple regression model statistics for each multiple-regression model. The S_2 sample was used for this regression analysis.

	Model 1	Model 2	Model 3	Model 4
Model-level statistics				
Multiple R^2	0.04	0.34	0.37	0.40
F statistic p value	<0.001	<0.001	<0.001	<0.001
Residual standard error	15.9	13.2	12.9	12.6
Predictors				
Intercept	b_k 57.5	b_k 18.0	b_k 20.7	b_k 40.6
	p value <0.01	p value <0.01	p value <0.01	p value <0.01
FEMALE	b_k -6.8	b_k -4.6	b_k -4.7	b_k -3.2
	p value <0.01	p value <0.01	p value <0.01	p value <0.01
BEMA Pre-test	b_k 0.32	b_k 0.32	b_k 0.33	b_k 0.30
Physics 1 GPA	b_k 6.7	b_k 6.9	b_k 6.9	b_k 6.9
Combined math score	b_k 2.1	b_k 2.1	b_k 1.8	b_k 0.92
CLASS pre-test	b_k 0.16	b_k 0.16	b_k 0.14	b_k 0.12
	p value <0.01	p value <0.01	p value <0.01	p value <0.01
Semester J2-1	b_k 1.7	b_k 1.7	b_k 1.7	b_k 1.5
Semester G2-1	b_k -7.1	b_k -7.1	b_k -7.1	b_k -6.8
Semester L2-1	b_k -4	b_k -4	b_k -4	b_k -3.8
Semester G2-2	b_k -1.4	b_k -1.4	b_k -1.4	b_k -1.2
Semester N2-1	b_k -0.48	b_k -0.48	b_k -0.48	b_k -0.67
	p value 0.28	p value 0.02	p value 0.34	p value 0.76
	p value <0.01	p value <0.01	p value <0.01	p value <0.01
Physics 1 Avg. Exam Score	b_k 8.0	b_k 8.0	b_k 8.0	b_k 8.0
	p value <0.01	p value <0.01	p value <0.01	p value <0.01

We also include variables controlling for the semester that each student took Physics 2 in Model 3. Using Model 3, we can estimate the difference between a male's and female's scores when controlling for prior physics course performance, prior math performance, and prior attitudes and beliefs to be 4.7 points. This is a smaller reduction in the gender gap than we saw using sample S_I . Controlling for these background factors, we account for about 30% of the observed gender gap in BEMA post-test scores.

Because the gender difference in Physics 1 course grade is small, we include an average Physics 1 exam score variable in lieu of Physics 1 grade in Model 4. We want to see if more of the gender gap can be accounted for by exam score, which has a larger gender gap than course grade. The average exam score is calculated by first converting each of the Physics 1 exam scores (three midterm exams and the final exam) to z -scores, and then computing the average exam z -score for each student in the sample. Converting to z -scores is a way to normalize the exam scores since each exam has a different average score. From Table 12, the average exam score is a significant predictor of BEMA post-test, and it reduces the *FEMALE* coefficient from -4.6 points to -3.2 points. When we use only the exam component of Physics 1 grade, we find that the resulting gender gap is only 3.2 points, approaching what we found with the S_I sample.

Again, in this sample, we included years of high school physics, students' declared major in Physics 2, ethnicity, and interaction variables between *FEMALE* and all other variables in the regression model. None of these variables significantly contributed to the model beyond those already included in the final model. This result is likely due to correlations between these variables and variables already included in the final model.

V. Discussion and Conclusions

In this study, we have examined in detail, three gender differences in the second-semester introductory physics course: retention, performance, and attitudes and beliefs. This has allowed us to expand our understanding of gender differences at our institution. We began by tracing the trajectories of students from Physics 1 to Physics 2. We found that, overall, males and females continued and did not continue from Physics 1 to Physics 2 at the same rate. This may be largely due to course requirements of engineering and science majors, most of whom are required to take both Physics 1 and Physics 2. However, we find differences when we focus on physics majors. While the gender differences in how many students did not take Physics 2 and how many students added and dropped the physics major are not significant, the differences are in a consistent direction such that the percentage of female physics majors in Physics 2 is significantly less than the percentage of male physics majors. We are disproportionately losing female physics majors as compared to male physics majors, an issue that needs to be further investigated.

Looking at performance in the second-semester course, we find that despite apparently equal pre-course E&M content exposure, males outperform females on the BEMA at the end of the semester. Though this may demonstrate bias in our courses, we argue that the BEMA pre-test does not accurately measure pre-course differences between males and females. And in fact, when we use the FMCE post-test as a pre-Physics 2 measure, we find that the gender gap may be reduced over the Physics 2 semester.

We also examine the course grades of males and females, as another measure of performance in the course. As we found in Physics 1, the total course grades of males and females are generally not different, as females outperform males on homework and participation,

but males outperform females on exams. This trend holds true for all semesters examined except two. In one semester, the gender differences on homework and participation were small, and males considerably outperformed females on the exams, resulting in significantly higher course grades for males. The other inconsistent semester was Semester N2-1, when females had significantly higher course grades than males. In this semester, there was no significant difference in the exam scores of males and females, and females considerably outperformed males on homework and participation. Semester N2-1 was also the only semester (in the past 25 semesters of Physics 1 and Physics 2 in which we have been collecting data) in which a female faculty member was the lecture instructor [35]. While the impact of a female faculty member on gender differences in the introductory physics courses needs to be further investigated, there is evidence that a female role model can influence the performance and interest of females in science and mathematics [33,36,37].

In addition to analyzing retention and performance, the third gender gap that we examined is in students' attitudes and beliefs. Just as in Physics 1, we find that both males and females shift toward less expert-like attitudes and beliefs over the course of Physics 2. However, the negative shifts that we observe in Physics 2 are between 0% and -6%. This is smaller than the shifts in Physics 1, which are typically between -5% and -15%. In all categories except one, Personal Interest, males and females do not have significantly different shifts. In the Personal Interest category, males have about a -2% shift while females have about a -5% shift. Because of the large pretest gender difference in the Personal Interest category and the significant gender difference in shifts, the Personal Interest category has the largest gender difference at the end of Physics 2, a difference of 11%. What was, at the beginning of Physics 1, an 8% gender difference in the Personal Interest category has increased to an 11% gender difference after just

two semesters of introductory physics. It seems that we are differentially negatively impacting females' interest in physics.

In trying to understand the possible sources of the gender disparities that we observe in E&M course performance, we used a multiple regression analysis to determine which factors contribute to students' post-test scores and could account for portions of the gender gap. We find that about 60% of the gender difference can be accounted for by differences in males' and females' prior conceptual performance on both the FMCE and the BEMA, prior math performance, and pre-course attitudes and beliefs about physics. That is, the gender gap in BEMA post-test scores is reduced from about 7% to about 3% when these measures of student background are controlled for. This result is the case when we use the FMCE post-test score as a measure of Physics 1 performance. We can instead use students' Physics 1 course grade, and when we do that, we find that less of the gender gap can be accounted for, only about 30%. If we use students' Physics 1 exam average in place of the FMCE post-test, we find that about 53% of the BEMA post-test gender gap can be accounted for. These differences in how much of the gender gap can be accounted for by different variables may suggest that the gender gap we observe is in part an issue of testing. The FMCE post-test and average exam grades (both tests) can account for a higher fraction of the BEMA post-test gender gap than can total course grade (made up of tests, homework, and participation). We observe repeated gender differences in performance on tests, which are high stakes, sequestered, time-sensitive tasks. These trends suggest that stereotype threat [38,39] may be playing a role in our courses, and affecting females' performance on tests, even tests that are explicitly used only for diagnostic purposes. The impact of stereotype threat and the alleviation of the threat through self-affirmation [40,41,42] will be discussed in Chapter 6.

From this work we can draw several conclusions. First, interactive engagement is not sufficient for eliminating, or even reducing, the gender gap. As suggested by our prior study, and further emphasized by this study, we need to explore the contextual factors in our classrooms that can impact the gender gap. By examining gender differences in Physics 2, we begin to investigate the impact of different contextual factors on the gender gap. Student familiarity with the course content may be an important factor in the gender gap, as is suggested by the smaller post-course gender differences in Physics 2, compared to Physics 1. However, students' familiarity with the current course content is not the only factor that contributes to course performance. Physics 1 performance is also a significant predictor of BEMA post-test score. Our prior study also suggested that the instructor may be a factor in the gender gap, as the post-test gender gap varied semester to semester. This is further supported by the current results, which also hint that the gender of the professor may play a role in the gender gap. These, and other, contextual factors need to be further investigated to determine if and how they influence gender disparities in the classroom.

We have seen from this work that differences in males' and females' backgrounds can account for much of the difference we observe at the end of the Physics 2 semester. This finding suggests that females are coming into our courses under-prepared, and leaving our courses under-prepared for future courses, as compared to males. In some sense, because the post-BEMA gender gap is smaller than the post-FMCE gender gap, we may say that females are catching up to the males. On the other hand, we may conclude that females are getting more and more behind males as they move through the introductory sequence, since they perform worse on tests of mechanics conceptual understanding and subsequently perform worse on tests of E&M conceptual understanding. Further, gender differences in students' personal interest in physics

seems to be increasing as students work through the introductory physics sequence. Females are more likely to leave the physics major than males. While none of these differences are particularly large on their own, females are consistently lagging behind males. Valian refers to this building up of deficiencies as an “accumulated disadvantage” [43]. Small, consistent differences can build up and accrue over time to result in large disparities.

Rather than identifying a single factor that is responsible for gender disparities in physics participation, we find small gender differences across several different factors, including retention, performance, and attitudes and beliefs. Female students consistently fall behind males in each of these areas as they move through the introductory physics sequence. This pattern of disadvantage suggests a systematic culture in which males are privileged over females. As mentioned in Chapter 2, Tatum refers to this cultural bias as a “smog of bias” [44], a smog that surrounds us and that we constantly breath in, though at times we may be unaware that it even exists. Understanding that retention, performance, and attitudes and beliefs are some of the mechanisms by which the cultural bias is maintained and reinforced is a first step towards alleviating the gender disparities in physics. By creating new cultural norms in our classrooms that are inclusive and supportive of all students [45], we may begin to construct physics classrooms and physics cultures in which males and females can participate equally.

Peer-reviewed publications based on this work

L. E. Kost-Smith, S. J. Pollock and N. D. Finkelstein, Gender disparities in second-semester college physics: The incremental effects of a “smog of bias”, *Physical Review Special Topics: Physics Education Research* **6**(2), 020112 (2010).

References (Chapter 3)

- [1] M. Lorenzo, C. Crouch and E. Mazur, Reducing the gender gap in the physics classroom, *American Journal of Physics* **74**, 118 (2006).
- [2] G. Carter, S. L. Westbrook and C. D. Thompkins, Examining science tools as mediators of students' learning about circuits, *Journal of Research in Science Teaching* **36**(1), 89 (1999).
- [3] J. B. Kahle, L. H. Parker, L. J. Rennie and D. Riley, Gender differences in science education: Building a model, *Educational Psychology* **28**(4), 379 (1993).
- [4] C. P. Russel and D. P. French, Factors affecting participation in traditional and inquiry-based laboratories, *Journal of College Science Teaching* **31**, 225 (2002).
- [5] J. Docktor and K. Heller, "Gender differences in both Force Concept Inventory and introductory physics performance" in *Proceedings of the 2008 Physics Education Research Conference*, edited by C. Henderson, M. Sabella and L. Hsu (AIP Conference Proceedings, New York, 2008) p. 15.
- [6] R. H. Tai and P. M. Sadler, Gender differences in introductory undergraduate physics performance: University physics versus college physics in the USA, *International Journal of Science Education* **23**, 1017 (2001).
- [7] Z. Hazari, R. H. Tai and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, *Science Education* **91**, 847 (2007).
- [8] P. M. Sadler and R. H. Tai, Success in introductory college physics: The role of high school preparation, *Science Education* **85**, 111 (2001).
- [9] D. E. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores, *American Journal of Physics* **70**, 1259 (2002).
- [10] E. Mazur, *Peer Instruction: A Users Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- [11] CAPA, <http://www.lon-capa.org/>; Mastering Physics, <http://www.masteringphysics.com/>.
- [12] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- [13] V. Otero, N. D. Finkelstein, R. McCray and S. Pollock, Who is responsible for preparing science teachers? *Science* **313**, 445 (2006).

- [14] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: implementing tutorials in introductory physics, *Physical Review Special Topics – Physics Education Research* **1**, 010101 (2005).
- [15] C. Turpen and N. D. Finkelstein, Not all interactive engagement is the same: Variations in physics professors' implementation of *Peer Instruction*, *Physical Review Special Topics – Physics Education Research* **5**, 020101 (2009).
- [16] In general, two professors are assigned to each of the large-enrollment introductory physics courses at CU. One professor (the lead) is responsible for the lecture part of the course and the other professor (the back-up) is responsible for the recitation part of the course. The back-up instructor usually does not have much, if any, interaction with students, while the lead instructor is who the students see in lecture. This being the case, we will only talk about the role of the lead instructor in these courses.
- [17] L. Ding, R. Chabay, B. Sherwood and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Physical Review Special Topics – Physics Education Research* **2**, 010105 (2006).
- [18] The pretest BEMA data for this course is reasonably stable (mean scores between 25% and 28%, standard deviations are about 9%) from semester to semester, which is why some faculty have chosen not to give the BEMA pretest to their students.
- [19] Data on students' high school experience (high school courses, high school GPA, SAT-math, and ACT-math scores), student demographic data (gender, ethnicity, and declared major), and course data (semester of enrollment and course grade) were collected from the Office of Planning, Budget, and Analysis at CU-Boulder. None of these data were self-reported.
- [20] While we recognize that there are other ways of quantifying a students' mathematics background, the most objective measures that were available to use were SAT- and ACT-Math scores and years of high school calculus. We found in the regression analyses that SAT- and ACT-Math scores were better predictors of student performance than years of high school calculus, so we chose to use those measures as a proxy for mathematics background.
- [21] For example, a student's SAT-math score S was converted to a z -score using the transformation $z_S = (S - \langle S \rangle) / SD$, where $\langle S \rangle$ is the average SAT-math score for all students and SD is the standard deviation of the SAT-math scores.
- [22] W. K. Adams, K. K. Perkins, N. Podolefsky, M. Dubson, N. D. Finkelstein and C. E. Wieman, A new instrument for measuring student beliefs about physics and learning physics: the Colorado Learning Attitudes about Science Survey, *Physical Review Special Topics – Physics Education Research* **2**, 010101 (2006).

- [23] The effect size for males is calculated as $ES_M = (\langle S \rangle_{Phys1\&2} - \langle S \rangle_{Phys1})/SD_M$, where $\langle S \rangle_{Phys1\&2}$ and $\langle S \rangle_{Phys1}$ are the average values for males who took both Physics 1 and 2, and for males who took Physics 1 only, respectively. SD_M is the standard deviation for all males who took Physics 1. A similar calculation is used for the effect size of the females.
- [24] The effect size is calculated as $ES = (\langle S \rangle_M - \langle S \rangle_F)/SD$, where the SD for all students is used.
- [25] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *American Journal of Physics* **66**, 64 (1998).
- [26] R. Chabay and B. Sherwood, Restructuring the introductory electricity and magnetism course, *American Journal of Physics* **74**, 329 (2006).
- [27] M. A. Kohlmyer, M. D. Caballero, R. Catrambone, R. W. Chabay, L. Ding, M. P. Haugan, M. J. Marr, B. A. Sherwood and M. F. Schatz, Tale of two curricula: The performance of 2000 students in introductory electromagnetism, *Physical Review Special Topics – Physics Education Research* **5**, 020105 (2009).
- [28] The normalized learning gain that we are using here is the gain for each individual student, calculated as $g = (post - pre)/(100 - pre)$.
- [29] The post-test scores for the FMCE and BEMA are slightly different. The mean FMCE post-test score is 65% and the mean BEMA post-test score is 56%. Even when we normalize the gender gaps by these post-test scores, the trends are the same.
- [30] At CU, students can drop a course within the first two and half weeks of the semester and the course will not appear on their transcript. If students drop the course after two and a half weeks, a grade of W (withdraw) will appear on their transcript. It is only these students, who receive a grade of W for the course, that are included in this analysis.
- [31] We presented here only average CLASS shifts over all Physics 2 semesters. While there are some differences in CLASS shifts by semester, most semesters show trends similar to the averages we've presented here. One exception is Semester D2-2, in which both males and females showed significant positive gains overall.
- [32] Z. Hazari, P. M. Sadler and R. H. Tai, Gender differences in the high school and affective experiences of introductory college physics students, *The Physics Teacher* **46**, 423 (2008).
- [33] S. E. Carrell, M. E. Page and J. E. West, Sex and science: How professor gender perpetuates the gender gap, NBER Working Paper 14959 (Cambridge, MA; May 2009). Available from <http://www.nber.org/papers/w14959>.

- [34] In Chapter 2, a similar multiple regression analysis on the FMCE posttest yielded an R^2 value of 0.44. Similar analyses conducted by other researchers have R^2 values of 0.16 (Ref. 6), 0.28 (Ref. 7), and 0.36 (Ref. 8). The R^2 value reported here of 0.34 is within this range.
- [35] Though not part of the data set used for this analysis, Professor N2-1 taught again in a more recent semester. We did not observe the same trends with respect to course grades or DFW rates the second time that she taught as the first time (details can be found in the Appendix). While we did not find consistent evidence of a positive impact of a female professor on the performance of females in the course, we still believe it is an issue that deserves further study.
- [36] D. M. Marx and J. S. Roman, Female role models: Protecting women's math test performance," *Personality and Social Psychology Bulletin* **28**, 1183 (2002).
- [37] J. G. Stout, N. Dasgupta, M. Husinger and M. A. McManus, STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM), *Journal of Personality and Social Psychology* **100**(2), 255 (2011).
- [38] C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance, *American Psychologist* **52**, 613 (1997).
- [39] S. J. Spencer, C. M. Steele and D. M. Quinn, Stereotype threat and women's math performance, *Journal of Experimental Social Psychology* **35**, 4 (1999).
- [40] C. M. Steele, "The psychology of self-affirmation: Sustaining the integrity of the self," in *Advances in Experimental Social Psychology*, edited by L. Berkowitz (Academic Press, New York, 1988) p. 261.
- [41] G. L. Cohen, J. Garcia, N. Apfel and A. Master, Reducing the racial achievement gap: A social-psychological intervention, *Science* **313**, 1307 (2006).
- [42] A. Martens, M. Johns, J. Greenberg and J. Schimel, Combating stereotype threat: The effect of self-affirmation on women's intellectual performance, *Journal of Experimental Social Psychology* **42**, 236 (2006).
- [43] V. Valian, *Why So Slow: The Advancement of Women* (MIT, Cambridge, 1999).
- [44] B. D. Tatum, *Why Are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race* (Basic Books, New York, 1997).
- [45] B. L. Whitten, S. R. Foster, M. L. Duncombe, P. E. Allen, P. Heron, H. M. Zorn, L. McCullough, K. A. Shaw and B. A. P. Taylor, What works? Increasing the participation of women in undergraduate physics, *Journal of Women and Minorities in Science and Engineering* **9**, 239 (2003).

CHAPTER 4: SELF-EFFICACY IN PHYSICS

I. Introduction

In this chapter, we explore students' sense of physics self-efficacy (defined in detail below) and the sources of self-efficacy, those experiences that contribute to the development of students' self-efficacy. We seek to answer the following research questions: 1) Are there gender differences in students' sense of physics self-efficacy and the sources of self-efficacy? 2) Does self-efficacy predict student performance in the course? 3) Does self-efficacy help account for the gender gap in performance that we observe, particularly once student background is already controlled? and 4) Do the hypothesized sources of self-efficacy predict students' end-of-semester self-efficacy, and are there gender differences in which sources are useful predictors?

We begin by outlining the prior literature on self-efficacy in science and the sources of self-efficacy in science. We also review the literature on gender differences in self-efficacy and work that has been done on self-efficacy in physics. We then summarize the research methods that were used in this study, including the development and administration of the Physics Self-Efficacy and Identity Survey, the sample of students analyzed, and the statistical methods that were employed. The results are presented in three major sections: students' sense of physics self-efficacy, sources of self-efficacy, and regression modeling. We find that males have significantly higher physics self-efficacy than females. These self-efficacy beliefs are useful predictors of students' FMCE and exam performance. For students in this study, background measures of physics and mathematics understanding account for about 70% of the gender gap in FMCE post-test scores. Students' self-efficacy beliefs account for an additional 12% of this gender gap. Also,

there are significant differences in student responses to questions about the four sources of self-efficacy, with males giving more positive interpretations of their experiences than females for each source. Each of the four sources of self-efficacy is a significant predictor of students' post-course self-efficacy in addition to students' pre-course self-efficacy and their performance in the course. Lastly, we discuss implications for the introductory physics courses and how they can be changed to positively impact students' self-efficacy.

II. Background

A. Self-Efficacy

Albert Bandura initially proposed self-efficacy as a way to account for different performance outcomes among individuals that were given similar treatments in his research on snake phobias. In one study [1], participants (all of whom had snake phobias severe enough to negatively impact their lives) were assigned to one of three conditions: 1) the Participant Modeling condition where participants were aided in interacting directly with a snake, 2) the Modeling condition, where the participant observed someone else interacting with the snake, or 3) the Control condition, where participants received no treatment. After being assigned to one of the three conditions, researchers assessed participants' self-efficacy related to interacting with snakes, and then participants were asked to complete tasks that involved increasingly threatening interaction with snakes. Bandura found a significant and positive relationship between participants' performance at the tasks and their self-efficacy, regardless of which treatment (or lack of treatment) that they received. The higher a participant's self-efficacy, the more tasks he/she successfully performed.

According to Bandura's social cognitive theory [2], self-efficacy is defined as, "people's judgments of their capabilities to organize and execute courses of action required to attain

designated types of performances” [p. 391]. In other words, self-efficacy characterizes a person’s confidence in or expectations about his/her ability to complete the tasks necessary for success at some activity. Bandura theorized that self-efficacy beliefs and expectations would influence whether people attempt an activity, how much effort they will put forth to achieve success, and how long they will persist when faced with obstacles, as well as performance at the activity. Self-efficacy beliefs differ from other expectancy beliefs in that they are task- and situation-specific and are related to a particular goal [3]. Lent and Brown emphasize that self-efficacy beliefs are dynamic and depend on the particular performance domain [4].

Though originally developed through research on phobias, Hackett and Betz [5] proposed using a self-efficacy approach to explain different career choices of men and women. They found that women had lower self-efficacy expectations compared to men of equal ability, with regard to completing the educational requirements and job duties of occupations that are traditionally held by men (e.g. accountant, mathematician, and engineer). Further, career self-efficacy was a significant predictor of participants’ interest in different occupations [6]. Researchers then began to focus on college science and engineering students to try to understand the lack of women in these fields. Self-efficacy beliefs were found to be useful predictors of both students’ choice of college major [7,8,9] and their perceived career options [10,11,12]. In one study, Hackett and Betz [13] found that mathematics self-efficacy was a better predictor of students choosing a mathematics-related major than were mathematics performance or achievement measures. Their results suggest that even though students’ self-efficacy is highly correlated with their past performance, both are useful predictors of future performance and career choices, and in some cases, students’ self-efficacy may be a better predictor.

In addition to major and career choice, self-efficacy has also proven useful in predicting mathematics and science achievement [3]. Self-efficacy has been shown to predict students' performance at a variety of tasks, including college science grades [10,11,14], mathematical problem-solving [15,16,17], and overall college performance [18]. In 1991, Multon, Brown and Lent [19] conducted a meta-analysis to analyze the relationship between self-efficacy and performance. Over 36 studies of self-efficacy, they found a significant and positive relationship between self-efficacy and student achievement (effect size of 0.38 [20]), where achievement was characterized as standardized test performance, course performance, or skill task performance (such as solving subtraction problems or reading comprehension). A similar result was also found in a smaller meta-analysis [21]. Self-efficacy has been shown to significantly correlate with student performance in math and science across a variety of contexts and performance assessments.

As hypothesized by Bandura [1], self-efficacy has also been found to predict persistence. Researchers have found that self-efficacy is related to students' persistence at completing specific tasks, like problem-solving [15], and student's persistence at a course of study, like the science major [10,11,14]. In their meta-analysis, Multon, Brown and Lent [19] found that over 18 studies, self-efficacy and persistence were significantly and positively correlated (effect size of 0.34 [20]). Self-efficacy has also been shown to correlate significantly with science and mathematics interest [9,21,22].

This prior work shows that self-efficacy is significantly related to students' choice of college major, perceived career options, performance, persistence, and interest in science and mathematics. Because of its significance to understanding students' choices and performance, we examine the physics self-efficacy of students in the introductory physics courses.

B. Sources of Self-Efficacy

In addition to proposing the construct of self-efficacy, Bandura also put forth that there are four sources of self-efficacy beliefs: *mastery experiences*, *vicarious experiences*, *verbal persuasion*, and *emotional responses* [1]. People interpret information from each of these four sources as they create and develop their self-efficacy beliefs. In this section I will describe each source of self-efficacy and then review the literature on how the sources influence self-efficacy beliefs.

Bandura hypothesized that *mastery experiences* would be the most powerful source of people's self-efficacy expectations [1,2]. Mastery experiences are experiences in which people engage in tasks similar or identical to the target task. For instance, if a student performed well on a physics exam, this would positively impact his/her self-efficacy regarding future physics exams. Similarly, experiences where people are not successful would negatively impact their self-efficacy. Because mastery experiences provide feedback to students on their own abilities on tasks related to the target task, it is thought to be the strongest source of self-efficacy beliefs.

When people do not have mastery experiences to draw on in assessing their ability to be successful at a target task, they can draw on *vicarious experiences*, where they observe other people. Observing others being successful or unsuccessful at a target task, can positively or negatively impact one's own self-efficacy. For someone who has never taken a physics course before, a peer's experience can be an important source of his/her beliefs about succeeding in physics. The most influential vicarious experiences are those in which the person being observed is similar in ability and demographics (e.g., age, gender, ethnicity, etc.) [23].

The third source of self-efficacy beliefs is *verbal* or *social persuasion*. This can take the form of encouragement or discouragement from trusted others, such as parents, teachers, or

peers. Words of support can increase a person's confidence in their abilities. It is important that the feedback be accurate and supportive, otherwise it could undermine peoples' self-efficacy [23,24].

The final source of self-efficacy is *emotional responses*. People interpret their physiological reactions while completing tasks as indications of their abilities. Students who feel anxious, nervous, or stressed while taking mathematics tests can interpret these emotions as evidence that they are not good at mathematics, regardless of performance or achievement. On the other hand, enjoyment, happiness, and a sense of calm can positively impact people's self-efficacy.

Researchers stress that it is people's *interpretations* of information from each of these four sources that is critical to developing self-efficacy, not the objective information itself [1,2,3,23]. Two students who both receive 75% on an exam had the same mastery experience, but they may perceive that experience differently so that it differentially contributes to their self-efficacy beliefs. Throughout this chapter, when we refer to mastery experiences, vicarious experiences, verbal persuasions or emotional responses, we mean students' *interpretations* of these experiences and not the experiences themselves.

There has been a significant amount of research into whether and to what extent these four sources contribute to self-efficacy. Findings are somewhat inconsistent, likely due to differences in populations and situations studied, as well as differences in research methodologies across studies [23]. A review by Usher and Pajares [23] gives ranges of correlation coefficients between each source and self-efficacy for over 30 studies. Correlations between self-efficacy and each of the sources have the following ranges: mastery experiences: 0.29 to 0.67; vicarious experiences: 0.09 to 0.58; verbal persuasion: -0.05 to 0.62; and emotional

responses: -0.57 to 0.41. One consistent result is that mastery experiences are a significant predictor of self-efficacy beliefs [25,26,27,28,29]. In their study of the relationship between the sources and mathematics self-efficacy, Lent, Lopez and Bieschke [30] found that although all four sources were significantly correlated with mathematics self-efficacy, only mastery experiences was a significant predictor of self-efficacy in a regression analysis. This was the case even when students' mathematics ACT score was included, suggesting that student performance and student perception of performance each make a unique contribution to self-efficacy.

Given the importance of self-efficacy to student performance and career choices, it is essential to understand how self-efficacy develops and what information students are drawing upon to evaluate their abilities. Additionally, these sources of self-efficacy provide potential levers for increasing students' physics self-efficacy, and thereby their physics performance. Indeed, studies have shown that interventions aimed at enhancing students' self-efficacy by providing mastery or vicarious experiences can result in higher self-efficacy [31,32]. In this chapter we explore not only students' sense of physics self-efficacy, but also how the introductory physics courses impact physics self-efficacy through each of the four sources.

C. Gender Differences in Self-Efficacy

As mentioned above, research into self-efficacy has long been motivated by differences between men and women. Hackett and Betz [5] originally proposed self-efficacy to explain higher concentrations of women in low-paying, low-status occupations and women's underrepresentation in mathematics, science and engineering occupations as well as in managerial and administrative positions. From there, several researchers have explored self-efficacy to explain the lack of women choosing to major in science and mathematics. Early researchers identified that females had significantly lower mathematics self-efficacy than males [7,12,16,25].

Regarding the sources of self-efficacy, the results are more mixed. Some researchers find no significant gender differences in the sources of self-efficacy [25,28], while others find that males have higher mastery experiences scores [12] or females have higher vicarious experiences and verbal persuasion scores [26]. In their study of middle school science students, Britner and Pajares [27] found that male and female middle school students did not have significantly different science self-efficacy. However, despite the female students earning higher grades in the class, the males reported higher mastery experience scores.

More recently, researchers have been interested in how the sources of self-efficacy differently predict self-efficacy for males and females. Again, the results are inconsistent. While some researchers report no gender differences in how the sources predict self-efficacy [25,26], others find that males and females may draw on different sources when developing their self-efficacy. Britner and Pajares [27] found that mastery experiences explained a greater percentage of the variance in self-efficacy scores for females than for males (35% versus 17%). In studying the academic self-efficacy of middle school students, Usher and Pajares [33] found that verbal persuasions contributed to self-efficacy more than mastery experiences for females, but was not a significant predictor of self-efficacy for males. In another study [34], they found that mastery experiences and emotional responses were significant predictors of academic self-efficacy for both males and females, but vicarious experiences were significant only for males and verbal persuasions were significant only for females.

In addition to these quantitative studies, there has been some qualitative work on how the sources of self-efficacy contribute to self-efficacy differently for males and females. In two related studies [35,36], Zeldin *et al.* analyzed the personal stories of males and females who selected and excel in science, technology, or math careers. They were looking to better

understand how self-efficacy influences the career choices of males and females. Researchers interviewed participants on their career development (how they selected their career) and history, as well as specific questions on each source of self-efficacy. They found that the critical sources of self-efficacy differed by gender. Females relied primarily on verbal persuasions and vicarious experiences in developing their self-efficacy beliefs, while males relied on their mastery experiences.

From this prior work it seems that self-efficacy and the sources of self-efficacy are important factors to consider in understanding gender differences in science and mathematics. The inconsistency in results suggests that there is still much to understand about how self-efficacy differs for males and females and how its development and creation varies by gender.

D. Self-Efficacy in Physics

Despite the large body of research on self-efficacy in science, relatively few studies have been done on self-efficacy in college physics. In this section I briefly review the work done on physics self-efficacy.

Most of the research done on self-efficacy in physics has looked at the impact of different pedagogies on students' self-efficacy. Cavallo, Potter, and Rozman [37] examined the self-efficacy beliefs of males and females at the beginning and end of an inquiry-based introductory physics course for life-science majors. They found that males had higher physics self-efficacy both pre- and post-course. Additionally, physics self-efficacy was a significant predictor of course achievement and conceptual understanding for both males and females. Fencil and Scheel [38] examined the impact of several different teaching strategies on the sources of self-efficacy. They found that active strategies, such as question and answer, collaborative learning, and conceptual problem assignments, which asked students to work together and engage with the

material, were the most effective at increasing students' physics self-efficacy. Modeling Instruction [39], a specific pedagogical approach where students collaborate in a studio-format class with integrated labs and lectures, was found to positively impact students' self-efficacy over traditional lecture [40]. Sawtelle, Brewer and Kramer [40] found that traditional lecture instruction negatively impacted all four sources of self-efficacy for both males and females. Modeling Instruction, however, resulted in no significant shifts in the sources of self-efficacy for males. Females in the Modeling course had significant positive shifts in the verbal persuasion source of self-efficacy, but the other three sources were not significantly impacted. In addition, pre-course self-efficacy was found to significantly predict the likelihood of students passing the course.

A second area of focus of the research on physics self-efficacy is its usefulness in predicting physics achievement. As stated above, some researchers report a significant and positive relationship between self-efficacy and course performance [40]. In a study of freshman physics in Turkey [41], researchers found that self-efficacy negatively contributed to physics course achievement, though not significantly. Shaw [42] examined the physics self-efficacy of males and females in three different introductory physics courses: conceptual physics, college physics, and university physics. Gender differences in self-efficacy were only significant in the conceptual physics course, where males had higher physics self-efficacy than females. In the college and university physics courses, males and females did not have significantly different self-efficacy. Shaw also examined the correlations between physics self-efficacy and course performance. For conceptual physics there were only weak correlations between self-efficacy and performance. For college physics, the correlation was significant only for females. And for university physics, the correlation was not statistically significant. Given the strong evidence for

self-efficacy as a predictor of performance detailed above, it is difficult to interpret these mixed results. It suggests that more research is needed to better understand the relationship between self-efficacy and performance in physics.

III. Research Methods

The goal of this study is to characterize the physics self-efficacy beliefs and the sources of self-efficacy of males and females in the introductory physics courses. We also want to understand how the sources of self-efficacy impact self-efficacy, and how self-efficacy then impacts student performance. Guided by the literature outlined in the previous section, a survey was developed to probe students' sense of physics self-efficacy and the sources of self-efficacy. In this section, we discuss the survey design and administration processes, the sample of students who responded to the survey, and the statistical methods that were used to analyze the survey results.

A. Physics Self-Efficacy and Identity Survey

The Physics Self-Efficacy and Identity Survey was designed to probe students' sense of physics self-efficacy, the sources of self-efficacy, and students' sense of physics identity (which will be discussed in detail in Chapter 5). These three areas comprised the three major sections of the survey. Here, we will discuss the design of the physics self-efficacy and sources of self-efficacy sections of the survey.

The survey was administered to students in Physics 1 and Physics 2 in three semesters. The course content and pedagogy of these semesters was in line with prior offerings of the course (as described in detail in Chapters 2 and 3). In the six semesters in which data were collected, four lecture instructors were involved. Throughout this chapter codes will be used to identify semesters and faculty (similar to codes used in prior chapters). The three semesters of

Physics 1 are identified as K1-1, J1-2 and K1-2. The three semesters of Physics 2 are identified as J2-2, N2-2 and G2-4. Three of the four instructors were male and one was female (N2-2). The survey was given to students twice during the semester. The pre-course survey was administered in the first or second week of the semester. The post-course survey was administered in the last few weeks of the semester. The survey was given online, and a link to the survey was either featured on the homepage of the course website or included as the last question on the weekly CAPA homework set. In most semesters, students had one week to fill out the online survey, though in semesters where the response rates were low after one week, the survey response time was extended by an additional week to allow more students to respond. Additionally, in these semesters where the response rates were low, an email was sent out to students who had not filled out the survey after one week to remind them to respond to the survey. In all semesters, the instructor offered students a “token” amount of extra credit (usually about equal to one homework problem) for taking the survey. Receiving the extra credit only required that students include their name when submitting the survey and did not require students to answer any questions.

Self-efficacy characterizes the beliefs that people have about their ability to complete the tasks necessary to achieve a particular goal. Self-efficacy beliefs are often assessed by asking people to rate their level of confidence in completing a specific task required for meeting a goal [3]. For students in introductory physics, the goal is to succeed in the physics course (we allow for success to mean earning a good grade in the course or learning physics, though presumably the two would be correlated). Based on the requirements of the course (doing well on homework and exams, participating in class, etc.) and the kinds of interactions that students are expected to engage in (discussing with other students, working in groups, asking questions, etc.), a list of

tasks required for succeeding in the physics course was generated. The survey asked students to rate how confident they were that they could complete each of the tasks *in the physics course*. Bandura, along with several other researchers, stress that measures of self-efficacy need to be specific to the tasks and situations that are required to achieve the particular goal that is being assessed in order for self-efficacy to be a useful predictor of performance [1,3,4,43,44]. Care was taken to develop tasks required for the introductory physics courses as they are taught at the University of Colorado, and to ensure that students report their confidence about completing the tasks in these particular courses. Researchers agreed that the list of tasks was accurate and comprehensive. On the survey instrument, larger, bold, italicized font was used to emphasize to students that they were to assess their confidence in the physics course.

In total, twenty-one tasks were generated, and students rated their confidence level on a Likert-like scale where the options were: Not Confident, A Little Confident, Somewhat Confident, Mostly Confident, and Very Confident. All of the survey questions can be found in Appendix A. This group of questions was asked both at the beginning and end of the semester on pre- and post-course surveys. On the pre-course survey, students were asked to rate their confidence level for completing the tasks in the current physics course in which they were enrolled (either Physics 1 or Physics 2). This assessment of pre-course self-efficacy was used to predict student performance in the course. On the post-survey, students were asked to rate their confidence level for completing the tasks in their next physics course (either Physics 2 or another physics course). If students were not planning to take another physics course, they were asked to rate their confidence level for completing these tasks in their next science course. By asking these questions on both the pre- and post-surveys, shifts in self-efficacy over the course of the physics class could be examined. Because of concerns about the length of the survey, only

fourteen of the twenty-one tasks were included in the surveys given in semesters K1-2 and G2-4. These fourteen tasks were chosen based on preliminary analysis of which tasks showed gender differences and were important for predicting performance in the course.

In addition to these multiple-choice items, there were two open-response questions that asked students how confident they were that they could 1) earn a B or better in the physics course and 2) learn the physics concepts. Students were also asked to explain their responses to each question. These questions were only asked on the surveys given in semesters K1-1, J1-2, J2-2 and N2-2. Student responses to these questions were not analyzed for this study.

We were also interested in the sources of self-efficacy. To assess how the introductory physics courses impacted students' self-efficacy through each of the four sources of self-efficacy, questions were adapted from the Sources of Self-Efficacy in Science Courses–Physics (SOSESC–P) instrument, developed by Fencil and Scheel [38]. The SOSESC–P instrument contains 33 statements to which students respond on a Likert-like scale. The items are distributed about equally among each of the four sources. Due to concerns about the length of the survey, we used only 15 of the 33 items from the SOSESC–P instrument. The items were chosen based on how relevant they were to our students, given the structure of our introductory physics course. The 15 sources of self-efficacy items are presented in Table 1.

Table 1. Survey items related to the sources of self-efficacy.

Sources of Self-Efficacy Items
Mastery Experiences
I received good grades on my homework in this class.
I was usually able to help my classmates with difficult physics problems.
I rarely knew the answer to the clicker questions raised in class.
I had difficulty with exams in this class.
Vicarious Experience
Watching other students in class made me think that I could succeed in physics.
Listening to other students during Tutorials made me think that I could <i>not</i> understand physics.
Students who were similar to me did well on exams.
Verbal Persuasion
My peers in this class encouraged me to do well on homework.
The instructor in this course encouraged me to put forth my best effort.
No one in this class has encouraged me to go on in science after this course.
People often tell me that I'm good at physics.
Emotional Responses
I was unable to think clearly when taking exams.
Physics makes me feel uneasy.
I got really stressed when working on homework.
I worried about my ability to solve problems on exams.
I enjoyed physics this semester.

Students responded to all fifteen of the sources of self-efficacy items by rating their level of agreement or disagreement from the following options: Strongly Disagree, Somewhat Disagree, Neutral, Somewhat Agree, or Strongly Agree. Of the fifteen items, five were asked on both the pre- and post-surveys, and the remaining ten were asked only on the post-survey. The five items that were assessed both at the beginning and end of the semester were related to the emotional responses source. Two slightly different versions of these items were given in the pre-course survey. In the K1-1 and J2-2 versions of the pre-course survey, students were asked to respond to the statements thinking about their experiences in prior *science* courses. In the pre-course survey given in the other four classes, students were asked to respond to the statements thinking about their experiences in a prior *physics* course (students could choose the response “I

have never taken a physics course” if they hadn’t taken physics before). In all versions of the post-course survey, students responded to the same statements reflecting back on their experiences in the current physics course (either Physics 1 or Physics 2). The remaining ten sources of self-efficacy items were asked only on the post-course survey, and students were asked to respond by reflecting back on their experiences in the current physics course. To summarize, five of the items related to the emotional responses source, four to the mastery experiences source, three to the vicarious experiences source, and three to the verbal persuasions source.

Additional demographic data (gender, declared major, and ethnicity) on students were collected from the Office of Planning, Budget, and Analysis at the University of Colorado, and student performance data (course grades and conceptual survey scores) were collected from the course instructors.

B. *Survey Sample*

In the six semesters in which the survey was administered, a total of 3216 students took the final exam in the course they were enrolled in, an indication that they completed the course. Of these, 2158 students took *both* the pre- and post-course survey, and so are included in the final survey sample. Students were considered to have taken the survey as long as they responded to at least one item. Survey response rates, broken out by course, semester, and gender are given in Table 2. Whole class survey response rates range from 54% to 79%, and the overall survey response rate (combining Physics 1 and Physics 2 over all semesters) is 67%. As mentioned in previous chapters, we find higher response rates among female students compared to male students.

Table 2. Survey response rates. The top half of the table gives the numbers of all, male and female students in Physics 1 and Physics 2 each semester included in this study. The bottom half gives the numbers and percentages of students who completed the pre- and post-course surveys in Physics 1 and Physics 2 each semester. Whole class survey response rates range from 54% to 79%.

	Entire Class			Took Survey		
	All	Males	Females	All	Males	Females
Physics 1	N	N	N	%	%	%
Semester K1-1	605	467	138	54	51	65
Semester J1-2	608	448	160	79	79	81
Semester K1-2	527	397	127	67	66	71
Total	1740	1312	425	67	65	73
Physics 2	N	N	N	%	%	%
Semester J2-2	524	398	126	68	62	87
Semester N2-2	458	357	101	66	64	70
Semester G2-4	494	392	99	68	65	84
Total	1476	1147	326	67	64	81

We also examine the demographics of students in the final survey sample. Table 3 gives the percentages of students in the final survey sample by gender, ethnicity, and declared major for each semester of Physics 1 and Physics 2. In most semesters, the demographics distributions of students who are in the final survey sample are significantly different from those students who are not in the final survey sample. This is important to keep in mind when generalizing results from the survey sample to the entire class population.

Table 3. Demographics of survey respondents. Frequencies for the gender, ethnicity, and student declared major for all students in the final survey sample. For each demographic variable, the distribution of students in the final survey sample was compared to the distribution of students not in the final sample using a χ^2 test. The χ^2 p value indicates whether the distributions in the two groups of students are significantly different. An underlined p value denotes that $p < 0.05$.

	Physics 1			Physics 2		
	Sem. K1-1	Sem. J1-2	Sem. K1-2	Sem. J2-2	Sem. N2-2	Sem. G2-4
Gender	%	%	%	%	%	%
Female	27.4	26.7	25.4	30.5	23.7	24.7
Male	72.6	73.3	74.6	69.5	76.3	75.3
χ^2 p value	<u>0.003</u>	0.67	0.36	<u><0.001</u>	0.25	<u><0.001</u>
Ethnicity	%	%	%	%	%	%
Asian	13.1	7.5	13.8	11.2	12.0	8.0
African American	1.2	1.7	1.4	2.0	1.3	0.6
Hispanic	3.7	6.6	5.6	5.0	4.0	7.1
Native American	0.3	0.4	1.1	0	0.3	1.5
White	72.9	75.8	74.0	73.4	72.3	75.6
Foreign	4.0	2.9	2.0	3.4	3.0	2.1
Unidentified	4.9	5.2	2.0	5.0	7.0	5.1
χ^2 p value	<u>0.023</u>	0.09	0.10	0.13	0.42	<u>0.016</u>
Declared Major	%	%	%	%	%	%
Physics	12.5	2.7	11.4	3.9	12.3	5.1
Engineering	44.8	59.0	47.8	69.2	46.3	63.7
Other Science	22.3	15.5	22.7	17.9	27.7	17.0
Non-Science	6.4	6.8	4.1	4.2	4.7	4.5
Undeclared	14.0	15.9	14.0	4.8	9.0	9.8
χ^2 p value	0.09	<u><0.001</u>	<u>0.004</u>	<u>0.026</u>	0.80	0.34

We also find significant differences in the distributions of these demographics variables between semesters, within Physics 1 and Physics 2. While the gender distributions are not significantly different by semester, the ethnicity distribution is significantly different by semester in Physics 1 and the distribution of declared majors is significantly different by semester in both Physics 1 and Physics 2. The differences in declared major are largely a result of physics majors, who almost exclusively take Physics 1 in the fall semesters and Physics 2 in the spring semesters. Additionally, as seen in prior chapters, there are gender differences in the distributions of student major and ethnicity among students who responded to the survey. It is important to

keep in mind when comparing results from one semester to another or when comparing males to females, that the samples of students likely differ along these demographic variables.

Because of the demographic differences that have been identified between survey respondents and non-respondents, and because survey respondents were self-selected, it is important to check for sampling bias. To do so, we compare the course grades of students who did and did not take the survey using a *t*-test (see Table 4 and Table 5). In all semesters of both courses, students who completed the survey had higher course grades than those who did not. The differences are statistically significant in all cases, except two semesters of Physics 2 in which the course grades of females who did and did not take the survey are not statistically different. This suggests that we are selectively sampling those students who are succeeding in the course, and our results may not capture the experiences of the lower performing students.

We also examine the course grade gender differences for students who are and are not included in the final survey sample. We find that in four of the six semesters, the gender gaps in the two groups are comparable. This suggests that gender differences we observe in the survey sample may be representative of gender differences in the larger populations. In the remaining two semesters, the gender gap in course grades among students who took the survey is much larger or smaller than the gender gap for students who did not take the survey. This suggests that in these semesters we may be over- or under-estimating the gender differences in the population by using only the final survey sample. These differences between the survey samples and the populations are important to keep in mind when interpreting the results.

Table 4. Comparison of course grades for students who did and did not take the survey in Physics 1. The * indicates that the difference between males and females or between survey respondents and non-respondents is statistically significant at the $p < 0.05$ level.

	Semester K1-1			Semester J1-2			Semester K1-2		
	M	F	M-F	M	F	M-F	M	F	M-F
Took survey (S)	2.89	2.79	0.10	2.83	2.60	0.23*	2.89	2.61	0.28*
Did not take survey (NS)	2.34	2.27	0.07	1.86	1.48	0.38	1.80	1.00	0.80*
S – NS	0.55*	0.52*		0.97*	1.12*		1.09*	1.61*	

Table 5. Comparison of course grades for students who did and did not take the survey in Physics 2. The * indicates that the difference between males and females or between survey respondents and non-respondents is statistically significant at the $p < 0.05$ level.

	Semester J2-2			Semester N2-2			Semester G2-4		
	M	F	M-F	M	F	M-F	M	F	M-F
Took survey (S)	2.84	2.38	0.46*	2.60	2.58	0.02	2.77	2.52	0.25*
Did not take survey (NS)	2.38	1.86	0.52*	2.19	2.04	0.15	2.25	2.27	-0.02
S – NS	0.46*	0.52		0.41*	0.54*		0.52*	0.25	

C. Statistical Methodology

Several statistical techniques were used throughout this chapter (and the next) to group variables and to make statistical comparisons between groups of students. Here I briefly discuss these techniques.

1. Factor Analysis

Recall that in assessing students' physics self-efficacy, a list of twenty-one tasks was presented to students and they were to rate how confident they were that they could complete each one. In order to organize and reduce this list of tasks, a factor analysis was used. Factor analysis is a statistical technique that reduces a large set of correlated variables into a smaller set of factors, which are linear combinations of the variables [45,46,47]. For this analysis, a

principal components analysis with a Varimax rotation was carried out. Varimax rotation is an orthogonal rotation method (meaning that the resulting factors are orthogonal), and it is the most popular rotation method. The factor analysis was carried out for all students, and for males and females separately. Additionally, the factor analysis was carried out on the pre- and post-course survey separately. Scree plots were examined to determine the appropriate number of factors. Tasks were included in a factor if the factor loading (the coefficient of each variable in the linear combination) was greater than 0.33. To ensure that the factors did not overlap, tasks that loaded on more than one factor at the 0.33 level were included in the factor with which it had the higher factor loading.

The factor analysis yielded slightly different factors for males and females (the factors for all students more closely aligned with the factors for males, since they make up the majority in the class), and for the pre- and post-course surveys (factors loadings can be found in Appendix B). Because of this, the factor analysis only informed the creation of categories of self-efficacy, and the factor loadings were not used in calculating the values of each category. We use the term category throughout the rest of the chapter to indicate that these groupings are not strict factors derived from a factor analysis. The final categories were created using only those tasks that fell into each factor for both males and females. There were a total of 18 tasks (of the 21 tasks total) that fell into one of four categories. Student responses to each task within a category were averaged to get a value for that category.

2. Statistical Comparisons

We are interested in not only examining the self-efficacy and sources of self-efficacy for males and females, but also in comparing males and females. When examining average values of student responses over several items, *t*-tests [48] were used to compare males to females. *t*-tests

are used to compare the average values of a variable between two independent samples. The data are assumed to be normally distributed and ratio level (continuous scale). The extension of the t -test, used when comparing more than two independent samples, is the One-Way Analysis of Variance (ANOVA) [49]. Similar to the t -test, ANOVA assumes normally distributed, ratio level data. For this analysis, ANOVA was used to compare male and female responses between the three semesters in which the survey was administered.

To compare male and female responses to individual items on the Physics Self-Efficacy and Identity Survey, Mann-Whitney U tests [50] were used. This parametric test is used to compare distributions from two independent samples, with no assumption that the data be normally distributed. The only requirement is that the data be ordinal level (rank-ordered) or higher. For most of the items on the survey, there is a clear rank order to the responses (for instance, Strongly Disagree < Somewhat Disagree < Neutral < Somewhat Agree < Strongly Agree). The Mann-Whitney U test is the analog of the t -test, but it used when the data is ordinal level, rather than ratio level. The extension of the Mann-Whitney U test (and the analog of the ANOVA), for comparing more than two samples is the Kruskal-Wallis test [50]. Similar to the Mann-Whitney U test, the Kruskal-Wallis test is used when the data is non-normal and ordinal level. When comparing student responses between the three semesters that the survey was administered, the Kruskal-Wallis test was used.

IV. Results

A. Students' Sense of Physics Self-Efficacy

1. Overall Self-Efficacy of Males and Females

We first examine the overall self-efficacy of males and females, to get a sense of students' beliefs about succeeding in all aspects of the course [51]. Students' overall self-efficacy

was computed by averaging their confidence ratings on all twenty-one of the course tasks that were listed on the survey. Because the list of tasks given on the survey in the K1-2 and G2-4 semesters was truncated, we do not include data from those semesters in this analysis of overall self-efficacy beliefs. Figure 1 presents the average overall self-efficacy of males and females at the beginning of the course for each semester in Physics 1 and Physics 2.

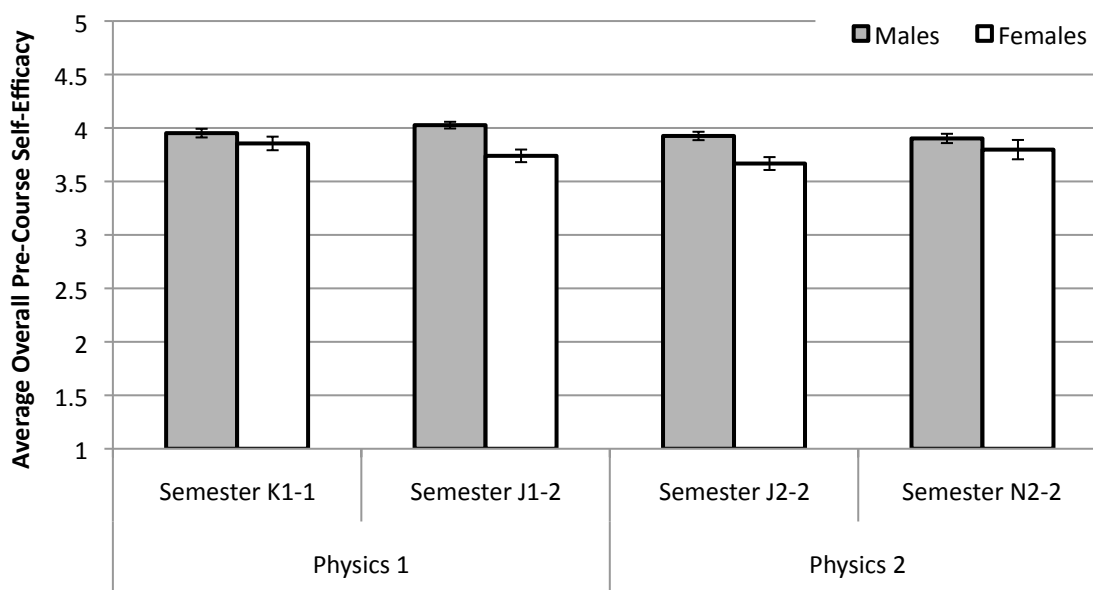


Figure 1. Average overall pre-course self-efficacy of males and females in Physics 1 and Physics 2 by semester. The overall self-efficacy score is the average of students' confidence ratings for all twenty-one tasks. Error bars represent the standard error on the mean. The differences between males' and females' overall self-efficacy are statistically significant in the J1-2 semester of Physics 1 and the J2-2 semester of Physics 2. In the remaining two semesters, the difference in self-efficacy between males and females is not statistically significant.

We find significant gender differences in students' overall pre-course self-efficacy, though not consistently. In Physics 1, there was not a significant difference between males' and females' self-efficacy in the K1-1 semester [$t(326)=1.26, p=0.21$], but the difference in the J1-2 semester was statistically significant [$t(479)=4.57, p<0.001$]. In Physics 2, there was a significant difference in the J2-2 semester [$t(355)=3.62, p<0.001$], but not in the N2-2 semester [$t(298)=1.13, p=0.26$]. We find significant gender differences when Professor J is teaching, but

not for the other two professors. Additionally, in the on-sequence semesters (the K1-1 semester for Physics 1 and the N2-2 semester for Physics 2), there are no significant pre-course self-efficacy gender differences. We only find significant differences in the off-sequence semesters. We cannot say from this data set whether it is the professor, the student population in the off-sequence semesters, both, or neither that is leading to the gender differences.

Despite the inconsistency in gender differences by semester, we don't find significant differences in the overall pre-course self-efficacy of males or females by semester [via ANOVA, all $p > 0.14$]. That is, the pre-course self-efficacy of males in Physics 1 does not vary by semester, nor does the pre-course self-efficacy of females in Physics 1. The same is true in Physics 2. This suggests that by gender, students come into the introductory physics courses with similar physics self-efficacy beliefs every semester. Additionally, students' average physics self-efficacy scores are between about 3.5 and 4 (which falls between Somewhat Confident and Mostly Confident), which suggests that students are fairly confident at the beginning of the semester about completing the tasks necessary for success in these courses.

The same list of tasks was given to students at the end of the course, and they were asked to rate their confidence in completing the tasks in their next physics or science course. This allows us to measure students' physics self-efficacy after completing either Physics 1 or Physics 2 (presumably, when they are better able to assess whether they can complete the tasks). Figure 2 presents the average overall post-course self-efficacy beliefs of males and females in each semester. Similar to the pre-course self-efficacy beliefs, we find inconsistencies when comparing males and females. In those semesters where there was a significant gender difference in pre-course self-efficacy, there is also a significant difference in post-course self-efficacy. In Physics 1, the post-course self-efficacy of males and females is not significantly different in the K1-1

semester [$t(325)=1.16, p=0.25$], but it is in the J1-2 semester [$t(477)=4.84, p<0.001$]. In Physics 2, the gender difference is significant in the J2-2 semester [$t(355)=5.50, p<0.001$], but it is not in the N2-2 semester [$t(297)=1.52, p=0.13$].

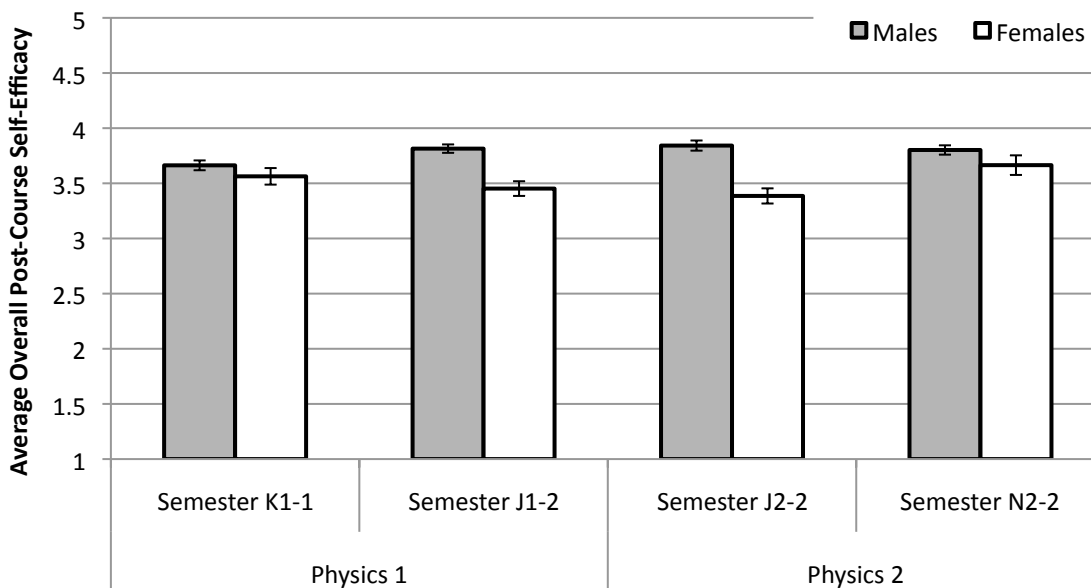


Figure 2. Average overall post-course self-efficacy of males and females in Physics 1 and Physics 2 by semester. The overall self-efficacy score is the average of students' confidence ratings for all twenty-one tasks. Error bars represent the standard error on the mean. The differences between males' and females' overall self-efficacy are statistically significant in the J1-2 semester of Physics 1 and the J2-2 semester of Physics 2. In the remaining two semesters, the difference in self-efficacy between males and females is not statistically significant.

We are also interested in how students' overall physics self-efficacy shifts over the semester. This gives some idea of how the course is impacting students' beliefs about succeeding in future courses. As is seen in Figure 3, all students shift significantly towards lower self-efficacy over the course of Physics 1 or Physics 2 [all $p<0.02$], except for females in the N2-2 semester of Physics 2 [$t(70)=1.64, p=0.11$]. The effect size of the shifts is between about -0.1 and -0.5, small to medium effect sizes. Most students become less confident about succeeding in the physics courses after taking Physics 1 or Physics 2. Further, this is equally true for males and

females in each semester. The gender differences in shifts in overall physics self-efficacy are not statistically significant in each semester [all $p > 0.07$].

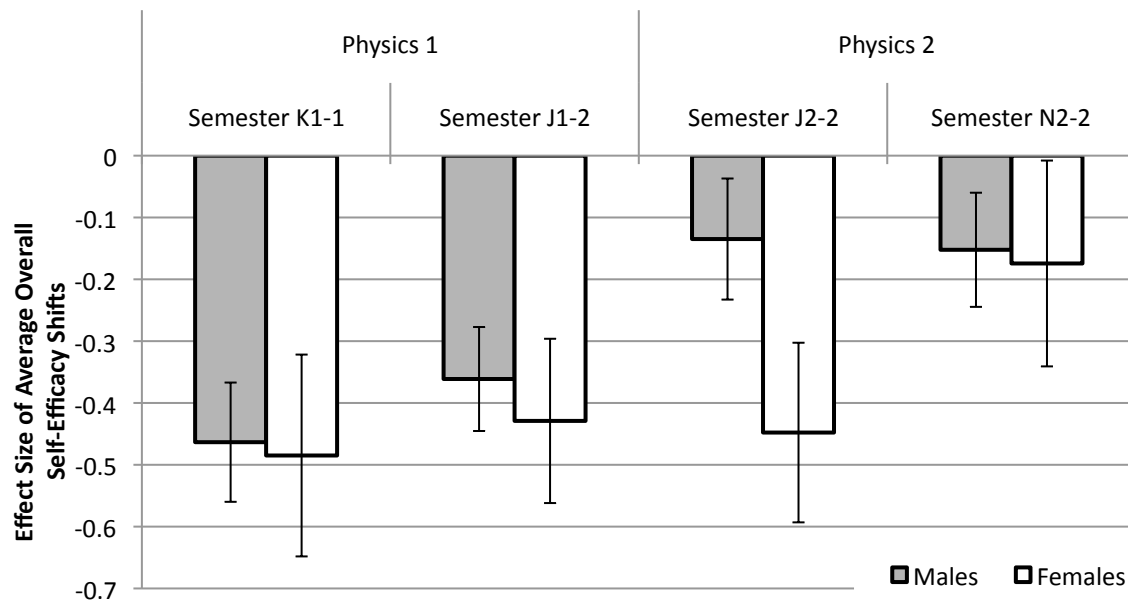


Figure 3. Effect size of shifts in males' and females' overall physics self-efficacy in Physics 1 and Physics 2. The effect size of the shifts is given by $ES = (post - pre) / SD_{pre}$. Error bars are given by $(SE_{pre}^2 + SE_{post}^2)^{1/2} / SD_{pre}$. All students shift towards lower overall physics self-efficacy, except for females in the N2-2 semester of Physics 2. Differences between males and females are not statistically significant in any semester.

2. Four Categories of Self-Efficacy

Factor analysis was used to divide the list of twenty-one tasks into categories or groups of tasks that were similar. By doing this, we could identify different dimensions of physics self-efficacy, and see whether students' confidence levels were different depending on the kinds of tasks. Additionally, we could explore whether males and females differed, or not, across all of the tasks, or if there were particular kinds of tasks for which gender differences were pronounced. The resultant factor loadings from the factor analysis for males and females are given in Appendix B. The final categories included the following tasks:

Performance

- Complete homework assignments by myself
- Perform well on exams
- Demonstrate what I know on exams
- Discuss physics ideas with my peers
- Defend my physics ideas to my peers
- Learn physics concepts
- Complete the course with a B or better
- Answer questions posed by the professor in class

Using Math in Physics

- Manipulate algebraic equations
- Use calculus to solve physics problems
- Interpret graphs

Getting Help

- Ask the professor for help when I get stuck on homework
- Ask a TA for help when I get stuck on homework
- Ask questions during lecture

Working with Other Students

- Complete homework assignments with the help of other students
- Find other students to help me when I get stuck on homework
- Find a group of students to study with
- Work well in a group during recitation

While the categories were determined using a statistical analysis based on student response patterns, the names of the categories were generated by the author based on a subjective judgment of how the tasks in each category were related to one another. Student confidence ratings for the tasks in each category were averaged to get a self-efficacy score for each student for each category. These category scores were then averaged for males and females in each semester that the survey was administered. Because of concerns about the length of the survey, tasks from the Getting Help category and the Working with Other Students category were not included on the survey given in semesters K1-2 and G2-4. Results for those categories include data only from the K1-1, J1-2, J2-2, and N2-2 semesters.

a. Performance self-efficacy

Figure 4 presents the pre-course *performance* self-efficacy for males and females in each semester of Physics 1 and Physics 2. In all semesters, females report significantly lower *performance* self-efficacy than males [all $p < 0.01$]. For tasks related to doing well on exams, discussing and defending their physics ideas, and doing well in the course, females are significantly less confident coming into the introductory physics courses than males.

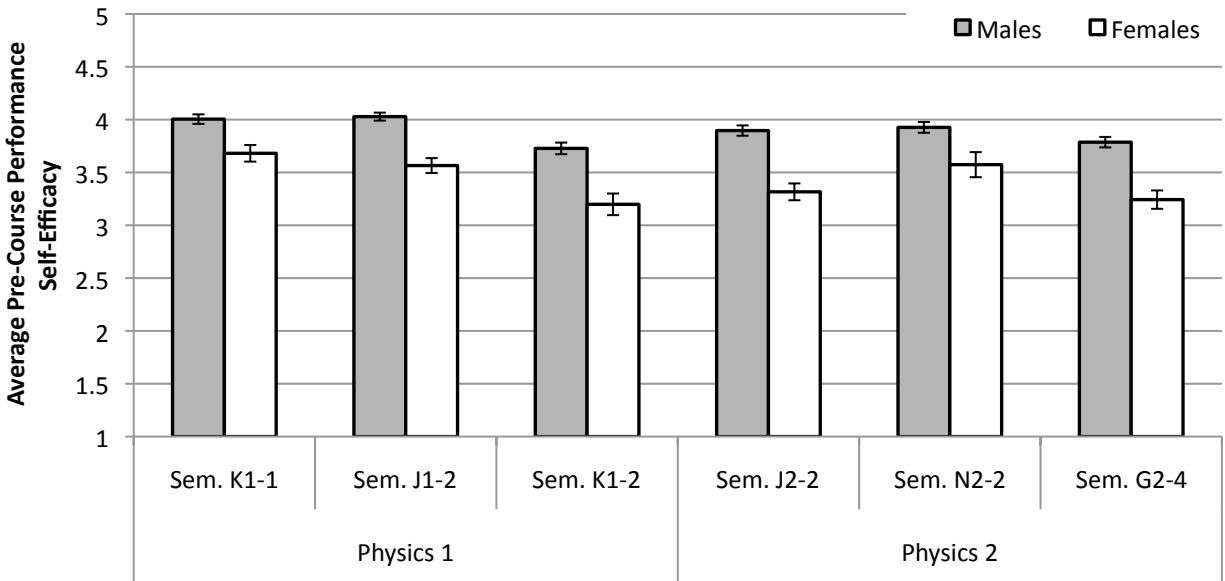


Figure 4. Average pre-course *performance* self-efficacy of males and females in Physics 1 and Physics 2 by semester. Error bars represent the standard error on the mean. The *performance* self-efficacy of females is significantly lower than the *performance* self-efficacy of males in all semesters.

Students in these introductory physics courses come from a variety of majors. It is possible that students from outside the physics or engineering major, or even outside the science major, would be less confident about succeeding in physics. Recall that the distribution of student majors in these courses differs by gender. A larger fraction of the females are other science and non-science majors compared to the males. The differences in student major distributions could explain the gender differences that are observed in physics *performance* self-efficacy. We can control for students' major by only comparing males and females who have the same declared major. Doing this, we find that in Physics 1, there are still statistically significant gender differences in *performance* self-efficacy for all student majors except physics (the difference between male and female physics majors is marginally significant, $p=0.09$). Accounting for student major does not explain the gender differences in *performance* self-efficacy in Physics 1, since these differences exist when comparing students within majors.

There are fewer gender differences in Physics 2, however, when student major is controlled. Gender differences in *performance* self-efficacy are only statistically significant for engineering and other science majors. In Physics 2, it appears that some of the gender gap in *performance* self-efficacy can be explained by differences in the distributions of males' and females' declared majors.

Gender differences in *performance* self-efficacy persist through the course. In all semesters, the gender differences in post-course *performance* self-efficacy are significant (all $p < 0.01$, see Figure 5). Again, we can control for student major, but the gender differences are significant in all majors except for the undeclared students in Physics 1. In Physics 2, the gender differences are only significant for the engineering and other science majors, as was seen in the pre-course *performance* self-efficacy scores. As with the pre-course *performance* self-efficacy, accounting for student major does not completely explain the gender differences that are observed.

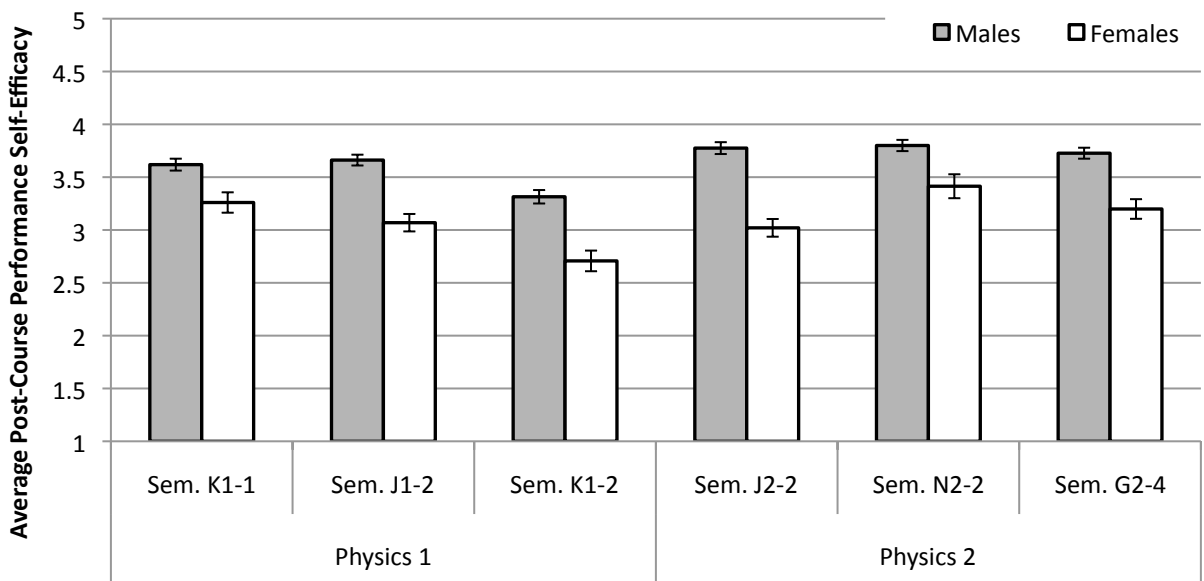


Figure 5. Average post-course *performance* self-efficacy of males and females in Physics 1 and Physics 2 by semester. Error bars represent the standard error on the mean. The *performance* self-efficacy of females is significantly lower than the *performance* self-efficacy of males in all semesters.

It is perhaps not too surprising that females are less confident about tasks related to exams and performance in the course, since we know from our prior work that they score significantly lower than males on exams. Again, we would like to control for the difference in actual exam performance between males and females, and then see if the difference in *performance* self-efficacy persists. We group students according to their average exam scores into quintiles [52]. We can then compare male and female average post-course *performance* self-efficacy within each quintile, and thereby only compare males and females who performed similarly on course exams. When we do this (see Figure 6), we find that except for students in the lowest exam performance quintile, females report lower post-course *performance* self-efficacy than their equally performing male counterparts. It is surprising that males and females who have the lowest exam scores have the same physics *performance* self-efficacy. Examining students in that quintile more closely, we find that females in that bin have significantly *lower* exam scores than males in that quintile, which makes the result even more puzzling. We do not have any explanation for why females in the lowest quintile do not have significantly lower *performance* self-efficacy than males in that bin, but it may be a question worth pursuing. Even controlling for the actual exam performance differences of males and females, a significant gender difference in *performance* self-efficacy exists for most students.

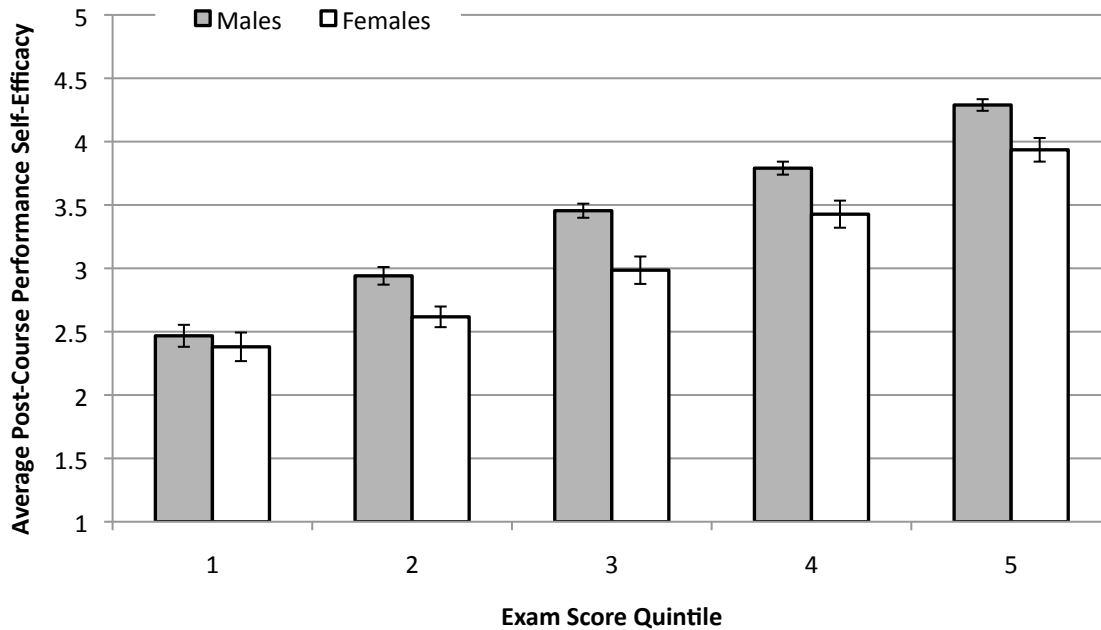


Figure 6. Males' and females' average post-course *performance* self-efficacy scores, where students are binned according to their course exam performance. In all quintiles except the lowest quintile (1), females have significantly lower post-course *performance* self-efficacy than males (all $p < 0.01$). Even when only comparing males and females with the same actual course performance, females still report less confidence about performing in the future.

As with overall physics self-efficacy, students' *performance* self-efficacy decreased over the course of Physics 1 for both males and females in all semesters [all $p < 0.01$] (see Figure 7). Males and females shifted equally in all semesters by about half of a standard deviation. In Physics 2, however, the results were less consistent. In the J2-2 semester, both males and females had significantly negative shifts in *performance* self-efficacy [males: $t(247) = 2.74$, $p < 0.01$; females: $t(108) = 4.69$, $p < 0.01$]. In the N2-2 semester, only males had a significant shift [males: $t(227) = 2.74$, $p < 0.01$; females: $t(70) = 1.68$, $p = 0.10$]. In the G2-4 semester, neither the males nor females' *performance* self-efficacy shifted significantly over the semester [males: $t(249) = 1.57$, $p = 0.12$; females: $t(81) = 0.40$, $p = 0.69$]. Though encouraging that the *performance* self-efficacy of males and females in the G2-4 semester did not decrease, they had the lowest pre-course

performance self-efficacy, and so ended the semester with post-course *performance* self-efficacy scores about equal to scores in the other semesters.

It also appears from Figure 7 that the shifts in *performance* self-efficacy are much smaller in Physics 2 compared to Physics 1. This may be a selection effect, since it seems that the pre-course *performance* self-efficacy of students in Physics 1 compared to Physics 2 is not different.

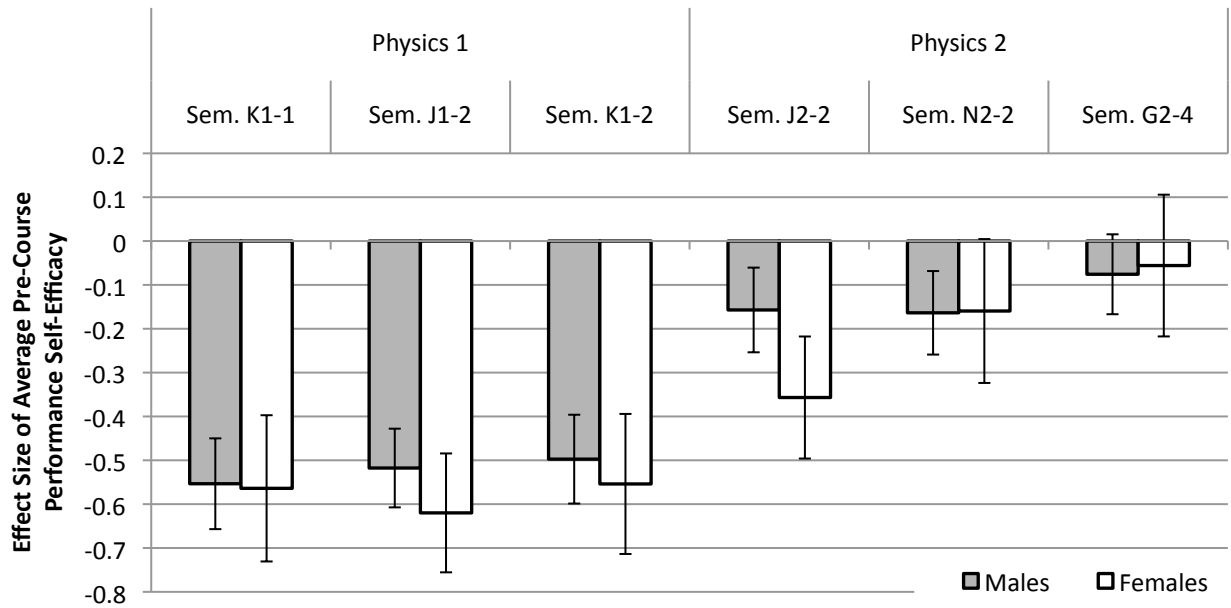


Figure 7. Effect size of shifts in males' and females' *performance* self-efficacy beliefs in Physics 1 and Physics 2. The effect size of the shifts is given by $ES = (post - pre) / SD_{pre}$. Error bars are given by $(SE_{pre}^2 + SE_{post}^2)^{1/2} / SD_{pre}$. All students in Physics 1 shift towards lower *performance* self-efficacy. Shifts for Physics 2 students were smaller and in some cases not statistically significant. Differences between males and females are not statistically significant in any semester, except the J2-2 semester of Physics 2.

b. Math in physics self-efficacy

As mentioned above, many studies have identified gender differences in students' mathematics self-efficacy, even among students pursuing science, engineering, or mathematics degrees. We would therefore expect to find gender differences in students' self-efficacy related to using mathematics in the introductory physics courses. In most semesters, this was not the case. At the beginning of the semester, there were no significant gender differences in students

math in physics self-efficacy, except in the J1-2 semester of Physics 1 [$t(478)=2.72, p<0.01$]. In all remaining semesters, males and females were equally confident about their ability to use mathematics in the physics course. Across all semesters, students' average pre-course *math in physics* self-efficacy was between about 4 and 4.25 (a rating of Mostly Confident), the highest average scores among all four of the self-efficacy dimensions. This is perhaps not unexpected in a calculus-based physics course for engineering and science majors.

The lack of gender differences in *math in physics* self-efficacy persists through the semester. On the post-course survey, we find that in four of the six semesters studied, there are not significant gender differences in students' confidence related to using mathematics in physics. In the J1-2 semester of Physics 1 and the J2-2 semester of Physics 2, females report significantly less confidence than males [$t(476)=2.72, p<0.01$ and $t(355)=3.03, p<0.01$, respectively]. Though we can identify individual semesters in which we observe a gender difference in students' mathematics self-efficacy, overall, the differences are small and mostly insignificant.

Shifts in students' *math in physics* self-efficacy are small (average effect size is -0.1), and in most semesters the shifts are not statistically different from zero. One exception is the J2-2 semester of Physics 2. Males in this semester had negative shifts of effect size -0.1 , not a significant shift [$t(246)=1.45, p=0.15$]. Females, however, had negative shifts of effect size -0.38 , a significant shift [$t(108)=3.98, p<0.01$]. This difference in shifts between males and females is statistically significant [$t(354)=2.70, p<0.01$].

c. Getting help self-efficacy

The third dimension of physics self-efficacy that was identified was self-efficacy related to getting help in the physics course from either the professor or a TA. Students' confidence in

these tasks was not assessed in the K1-2 and G2-4 semesters, so data from these semesters are not included in the following analysis. At the beginning of the semester, there are no significant gender differences in students' reported confidence about getting help in physics [all $p > 0.05$]. Average *getting help* self-efficacy scores are between about 3.3 and 3.8 (corresponding to between Somewhat Confident and Mostly Confident), and are the lowest average pre-course scores among all four of the physics self-efficacy dimensions. Compared to the other self-efficacy dimensions, students come into the introductory physics courses least confident about getting help from the professor and TAs.

Over the course of the semester some gender differences in *getting help* self-efficacy emerge. In two of the four semesters analyzed, females had significantly lower *getting help* self-efficacy compared to males. In the J1-2 semester of Physics 1 and in the J2-2 semester of Physics 2 females were less confident about finding help when they needed it [$t(476) = 2.78, p < 0.01$ and $t(355) = 3.32, p < 0.01$, respectively]. In the remaining two semesters, males and females did not have significantly different *getting help* self-efficacy beliefs [Physics 1: $t(325) = -0.67, p = 0.51$; Physics 2: $t(297) = -0.20, p = 0.84$].

Looking at shifts in students' *getting help* physics self-efficacy, all students had significantly negative shifts in their confidence about getting help in the physics course [all $p < 0.02$]. The shifts of males and females were not significantly different, except in the J2-2 semester of Physics 2, where males shifted by -0.14 standard deviations and females shifted by -0.44 standard deviations [$t(355) = 2.56, p = 0.01$]. Over the course of the semester, both males and females grew less confident in their ability to get help from the professor or the TAs.

d. Working with other students self-efficacy

The final physics self-efficacy dimension was the *working with other students* self-efficacy. Again, students' confidence in these tasks was not assessed in the K1-2 and G2-4 semesters, so data from these semesters are not included in the following analysis. In both the pre- and post-course surveys, there were no significant gender differences in students' *working with other students* self-efficacy. Males and females appear to be equally confident about working with other students in the course. Students' average *working with other students* self-efficacy scores were about 4 (corresponding to Mostly Confident). Also, the shifts in students' *working with other students* self-efficacy are not statistically different from zero in all cases, except for females in the J2-2 semester of Physics 2. Overall, there were very few gender differences related to students' confidence in working with other students, and students' confidence, which was relatively high, did not change much over the course of the semester.

3. Summary

Figure 8 presents the pre- and post-course scores for overall physics self-efficacy as well as each of the four dimensions of self-efficacy averaged over all semesters in which each dimension was assessed [53]. In the previous sections, we examined self-efficacy separately for each semester and found that the results were not always consistent. By looking at all of the students together, we are able to pick out trends in the data that may have been more difficult to parse when looking at results for individual semesters. When averaging over all semesters, we find that there are significant gender differences in students' overall physics self-efficacy both at the beginning and end of the semester [pre: $t(1464)=5.49$, $p<0.01$; post: $t(1460)=6.91$, $p<0.01$]. Males come in to and leave the physics courses more confident than females about their ability to complete the tasks that are necessary for success in physics. When we break the list of tasks

down into groups of similar tasks, we see that there are some areas in which males and females are equally confident. For instance, at the beginning of the semester, males and females are equally confident about using *mathematics in the physics* course [$t(2099)=1.87, p=0.08$], about *getting help* from professors and TAs [$t(1464)=0.74, p=0.46$] and about *working with other students* [$t(1464)=0.06, p=0.95$]. It is only the pre-course *performance* self-efficacy of males and females that is significantly different [$t(2101)=11.93, p<0.01$].

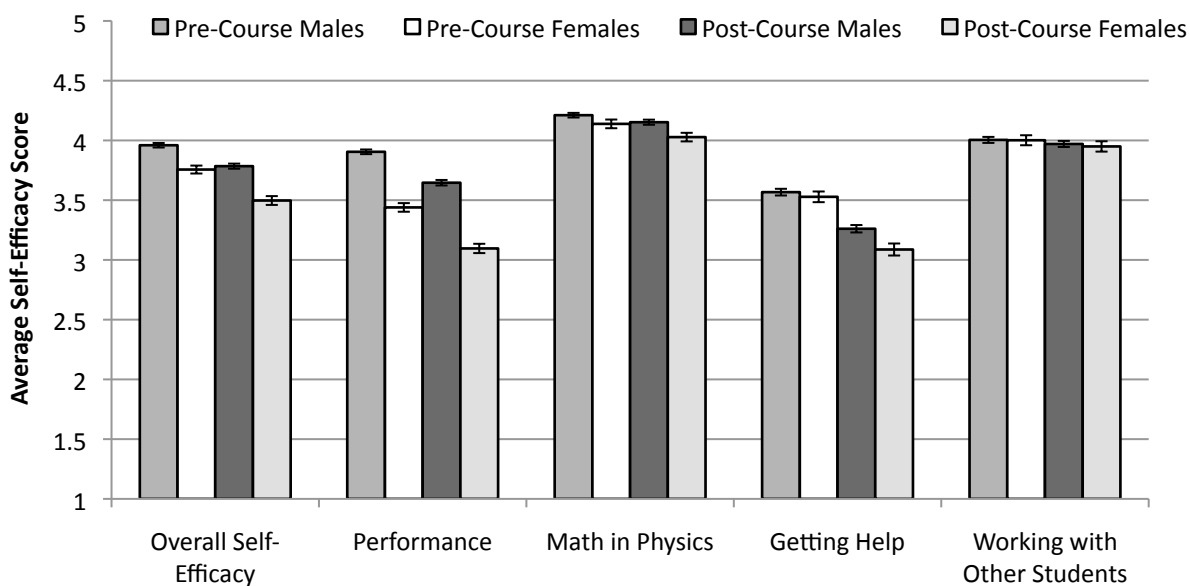


Figure 8. Average male and female, pre- and post-course scores for overall physics self-efficacy and for each of the four dimensions of self-efficacy. Scores were averaged over all semesters in which the dimension was assessed. Error bars represent the standard error on the mean.

Despite males and females being about equally confident in their abilities at the beginning of the semester, females are significantly less confident than males in three of the four dimensions of self-efficacy at the end of the semester. Females report lower self-efficacy related to their future physics or science courses in the *performance* [$t(2142)=12.24, p<0.01$], *math in physics* [$t(2139)=2.99, p<0.01$], and *getting help* [$t(1459)=2.93, p<0.01$] categories. These results suggest that males and females are having different experiences in the introductory physics

courses, which is leading to a larger decrease in self-efficacy for females. In the next section, we explore how males and females are interpreting their experiences, and thereby developing their self-efficacy, in the introductory physics courses by examining the four sources of self-efficacy beliefs.

B. *Sources of Self-Efficacy*

Bandura posited four sources of self-efficacy beliefs: mastery experiences, vicarious experience, verbal persuasion, and emotional responses. In order to better understand the gender differences in self-efficacy, which we identified in the previous section, we probed these four sources of self-efficacy to see if males and females were having experiences that differently contributed to the development of their self-efficacy. Because we were interested in what experiences they had in Physics 1 or Physics 2 that contributed to their self-efficacy and how their self-efficacy changed over the semester, most of the questions related to the sources of self-efficacy were only asked on the post-course survey. The one exception is questions related to the *emotional responses* source. These questions were asked on both the pre- and post-course surveys so that we could see how students' experiences in their prior science or physics course compared to their experiences in Physics 1 or Physics 2. Though the four kinds of experiences are labeled "sources" (in line with prior literature), we make no causal claims between students' experiences and their self-efficacy. We are not asking students whether these experiences led to changes in their confidence. Rather we are asking to what extent students had each of these kinds of experiences and how they interpreted the experiences as either positive or negative.

1. Mastery Experiences

The first of the four sources of self-efficacy, and what Bandura proposed would be the most powerful source, is *mastery experiences*. These are students' interpretations of their

successes and failures in the introductory physics course. Students were asked about their performance on homework, exams, difficult physics problems, and clicker questions (see Table 1). As presented in Figure 9, females report less positive interpretations of their *mastery experiences* than males in all semesters of Physics 1 and Physics 2 [all p 's<0.01]. Again, controlling for actual exam performance (as in Figure 6) does not fully explain the observed differences in reported *mastery experiences*. Females with exam scores in the top three quintiles of the class, still report less positive interpretations of their *mastery experiences* compared to equally performing males.

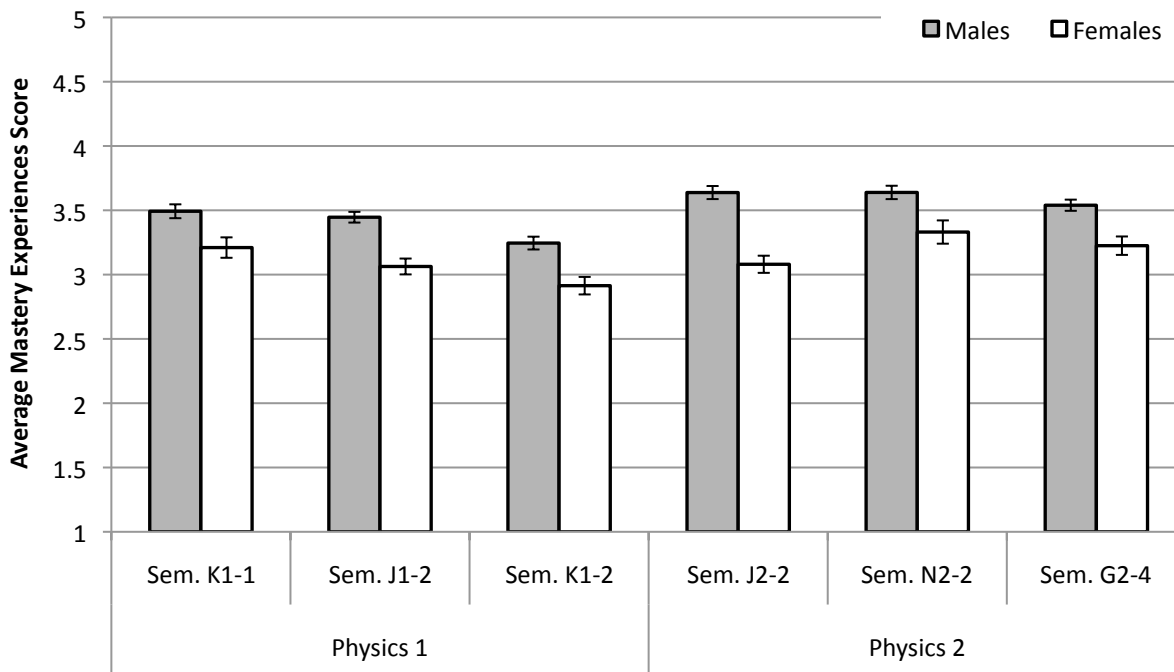


Figure 9. Average *mastery experiences* score for males and females in Physics 1 and Physics 2 by semester. Error bars represent the standard error on the mean. Females have significantly lower average *mastery experiences* scores than males in all semesters.

2. Vicarious Experiences

The second source of self-efficacy is *vicarious experiences*, or opportunities for students to observe others being successful or unsuccessful in the physics course. To assess the *vicarious*

experiences source, students were asked to what extent watching and listening to other students made them think that they could succeed in physics (see Table 1). Males' and females' average responses for each semester are presented in Figure 10. In all semesters of Physics 1 and Physics 2, females report less positive interpretations of their *vicarious experiences* than males [all p 's<0.05].

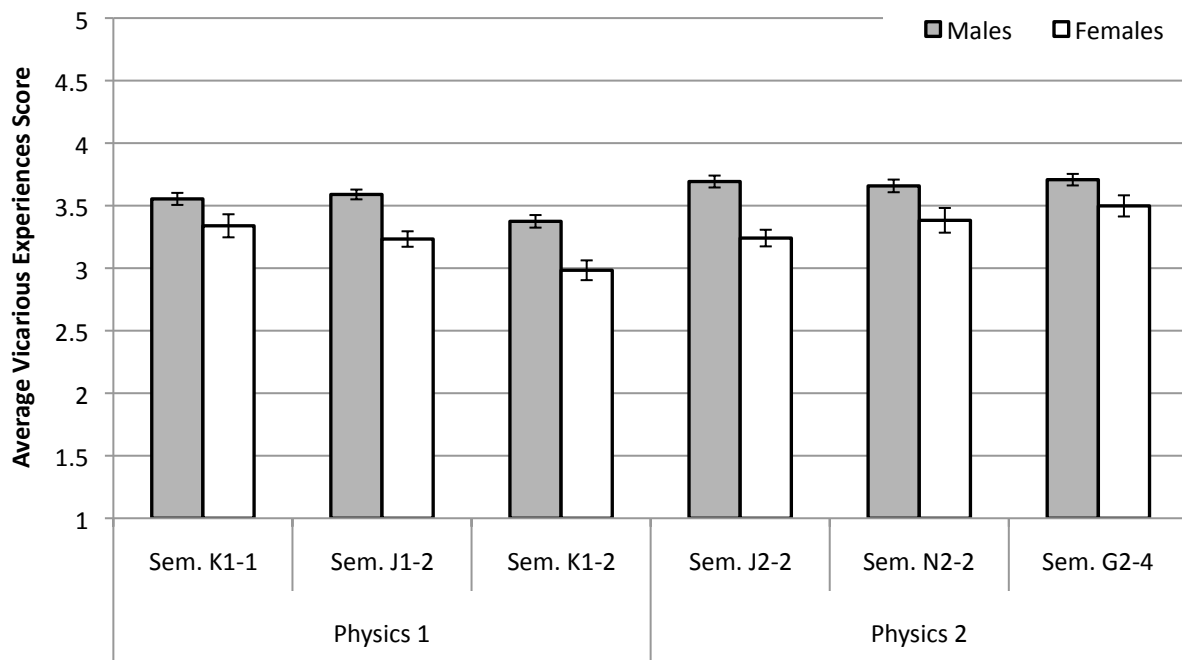


Figure 10. Average *vicarious experiences* score for males and females in Physics 1 and Physics 2 by semester. Error bars represent the standard error on the mean. Females have significantly lower average *vicarious experiences* scores than males in all semesters.

3. Verbal Persuasion

The third source of self-efficacy beliefs is *verbal persuasion*. The *verbal persuasion* source often takes the form of encouraging or discouraging messages from other people. We assessed this source with questions that asked to what extent students received encouragement from peers, the instructor, or any one else (see Table 1). In four of the six semesters of Physics 1 and Physics 2 that we analyzed, there is no significant difference in the amount of *verbal*

persuasion that males and females report receiving [all p 's>0.05] (see Figure 11). For the most part, males and females report receiving the same amount of encouragement in these courses. In the K1-2 semester of Physics 1 and in the J2-2 semester of Physics 2, females report significantly less *verbal persuasion* than males [$t(351)=3.13, p<0.01$ and $t(353)=3.69, p<0.01$, respectively]. Recall that these are students' interpretations of the verbal persuasion that they received, and therefore may not reflect the actual encouragement or discouragement that others may have given them. It is encouraging to see that for the most part males and females perceive that they are receiving the same level of encouragement to do well in these courses and to go on in science.

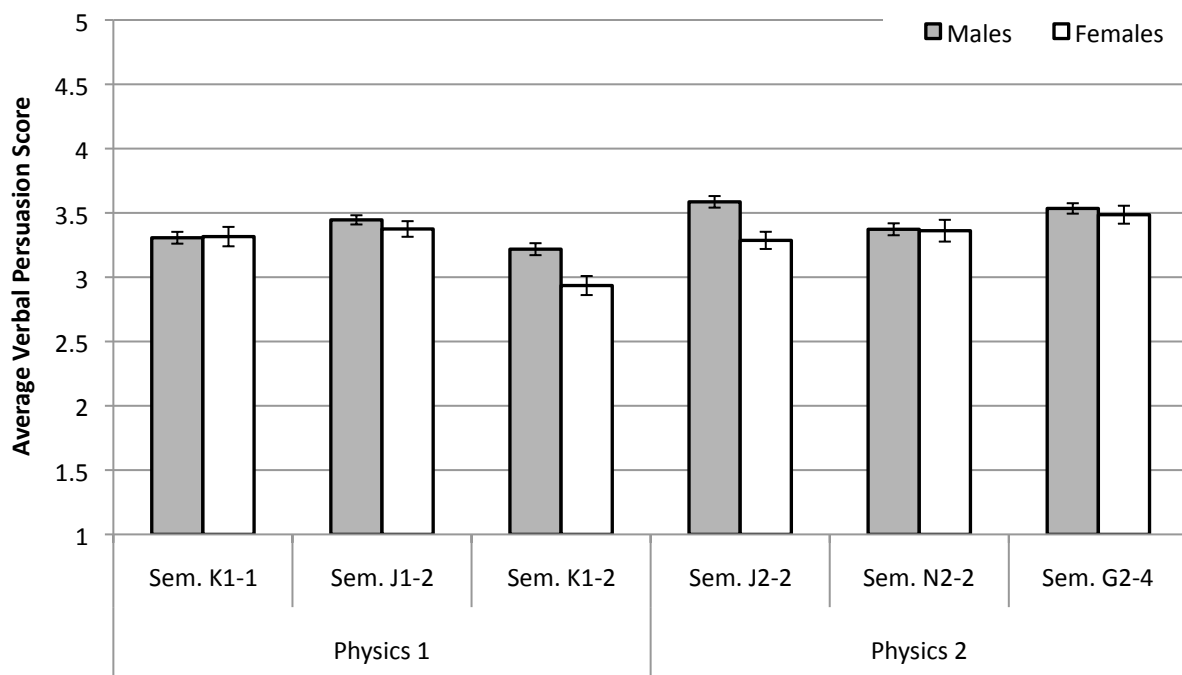


Figure 11. Average *verbal persuasion* score for males and females in Physics 1 and Physics 2 by semester. Error bars represent the standard error on the mean. Females and males do not have significantly different average *verbal persuasion* scores in all semesters, except K1-2 and J2-2.

4. Emotional Response

The final source of self-efficacy beliefs is *emotional responses*. These are students' interpretations of the physiological responses that they experience while in the physics course, doing homework or taking exams. Survey questions asked to what extent students experienced emotions such as stress, uneasiness, worry, and enjoyment (see Table 1). The *emotional response* questions were asked both in the pre- and post-course surveys. In the pre-course survey, students were asked about their experiences in a prior science (K1-1 and J2-2 semesters) or physics (J1-2, K1-2, N2-2, and G2-4 semesters) course. In all semesters of Physics 1 and Physics 2, females report more negative *emotional responses* in their prior science or physics course than males [all p 's<0.04] (see Figure 12). Females are more likely to report that they were worried, stressed, uneasy, and unable to think clearly in their prior science or physics courses than males [54].

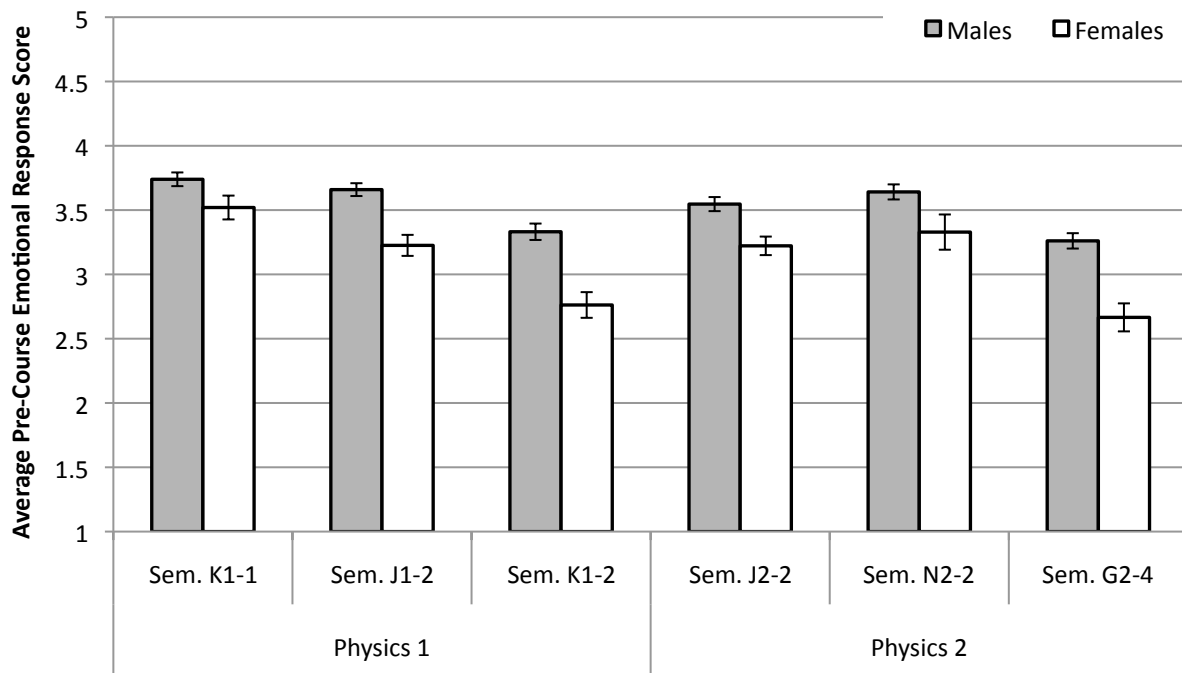


Figure 12. Average pre-course *emotional response* score for males and females in Physics 1 and Physics 2 by semester. Error bars represent the standard error on the mean. Students were asked about their emotional responses to doing homework and taking exams in their prior science (K1-1 and J2-2 semesters) or physics courses (J1-2, K1-2, N2-2, and G2-4 semesters). Females have significantly lower average *emotional response* scores than males in all semesters.

On the post-course survey, students were asked about their *emotional responses* in the Physics 1 or Physics 2 course. Similar to students' prior experiences, females continue to report more negative *emotional responses* than males in all semesters of Physics 1 and Physics 2 [all p 's<0.01] (see Figure 13). Even when controlling for actual exam performance or course grade (as in Figure 6), females continue to report more negative *emotional responses* than males.

Both males and females in the K1-2 semester of Physics 1 had unusually low *emotional responses* scores. This may be partly due to the unusually low exam scores in that semester. The average exam score in the K1-2 semester was 59%, compared to 69% and 66% in the K1-1 and J1-2 semesters.

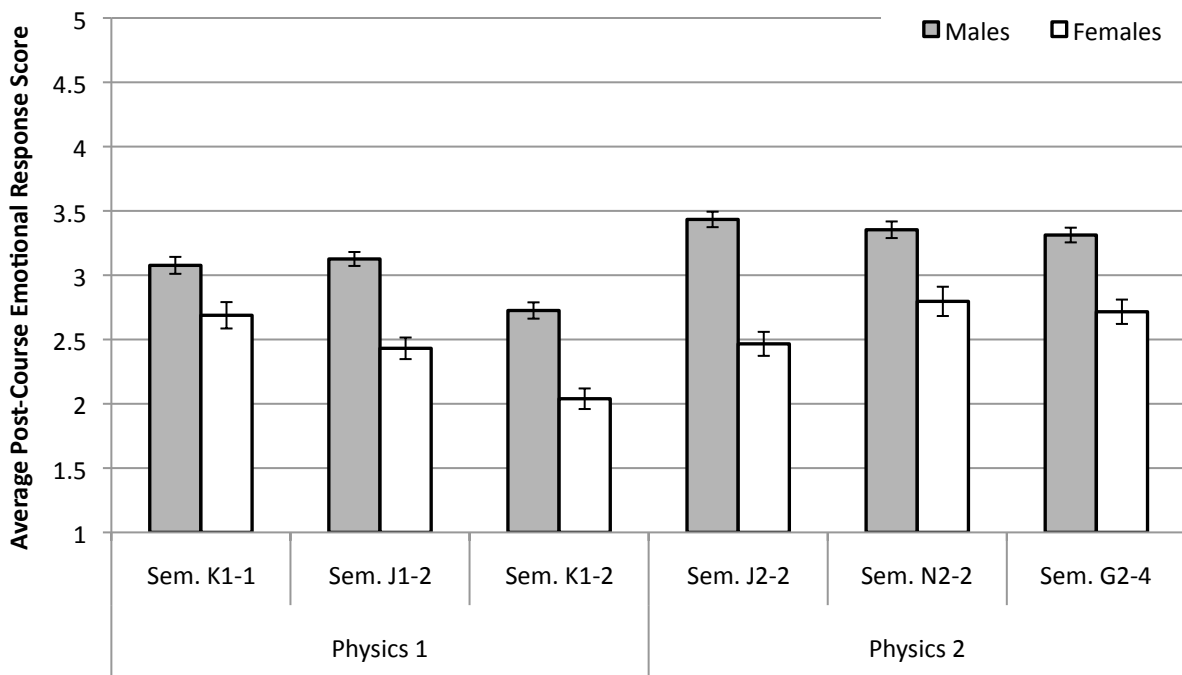


Figure 13. Average post-course *emotional response* score for males and females in Physics 1 and Physics 2 by semester. Error bars represent the standard error on the mean. Students were asked about their emotional responses to doing homework and taking exams in their current physics course. Females have significantly lower average *emotional response* scores than males in all semesters.

Because questions related to students' *emotional responses* were asked on both the pre- and post-course surveys, we are able to examine shifts in students' responses. This allows us to

compare students' reported *emotional responses* in prior courses to their emotional responses in their current Physics 1 or Physics 2 course. Shifts in students' *emotional response* scores are presented in Figure 14. Both males and females significantly shift towards more negative *emotional responses* in all semesters except the G2-4 semester of Physics 2. These negative shifts are especially large in Physics 1, where the effect sizes of the shifts are between -0.6 and -0.8. This suggests that students are more stressed and worried in the college introductory physics courses than in their previous science or physics courses. The gender difference in shifts is only statistically significant in the J1-2 semester of Physics 1 and the J2-2 semester of Physics 2. For the most part, the increases in stress and uneasiness are equal for males and females.

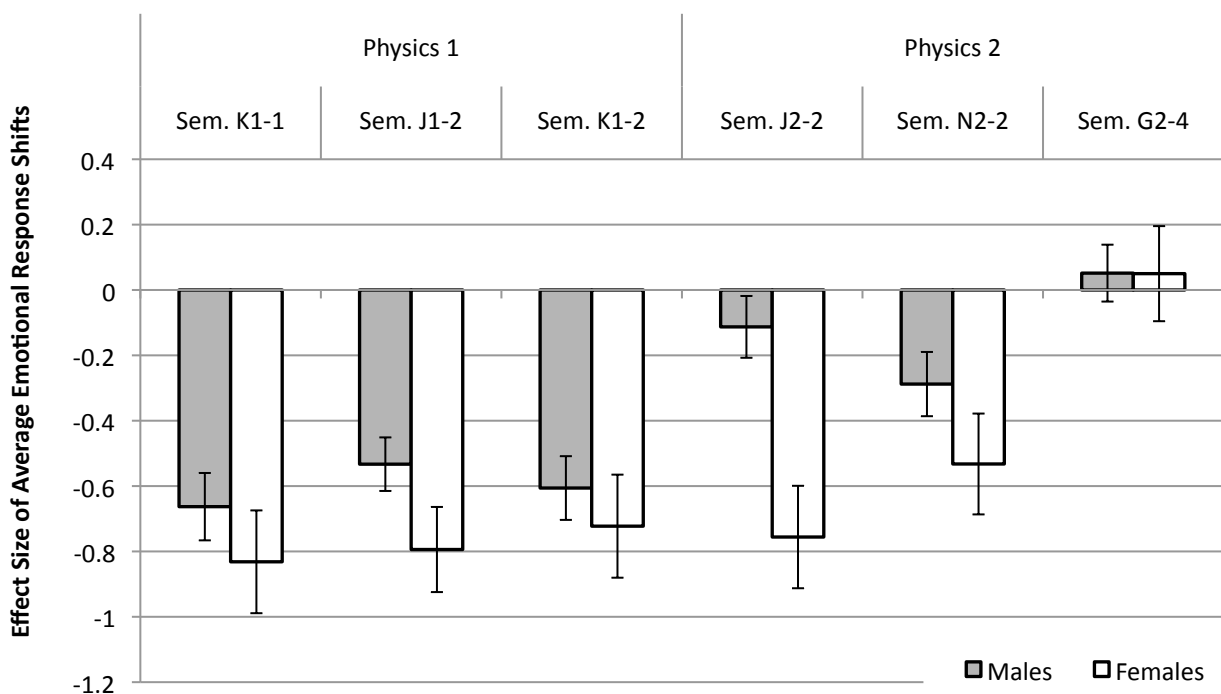


Figure 14. Effect size of the average shifts in the *emotional response* scores for males and females in Physics 1 and Physics 2 by semester. The effect size of the shifts is given by $ES = (post - pre) / SD_{pre}$. Error bars are given by $(SE_{pre}^2 + SE_{post}^2)^{1/2} / SD_{pre}$. Both males and females significantly shift towards more negative *emotional responses* in all semesters except the G2-4 semester of Physics 2. The gender difference in shifts is only statistically significant in the J1-2 semester of Physics 1 and the J2-2 semester of Physics 2.

5. Summary

In this section we have examined each of the four sources of self-efficacy and how males and females report their interpretations of these experiences. Figure 15 presents males' and females' average sources of self-efficacy scores for all students included in the analysis [53]. When averaging over all semesters, we find that females report less positive interpretations of each source of self-efficacy than males. The difference between males and females has an effect size of 0.48 for *mastery experiences*, 0.42 for *vicarious experiences*, 0.17 for *verbal persuasion*, and 0.66 for *emotional response*. The smallest difference is the *verbal persuasion* source suggesting, as was mentioned above, that males and females report that they receive equal amounts of encouragement from peers and instructors in the course. The differences in the remaining three sources are medium to large effects. Females are reporting less positive *mastery* and *vicarious experiences*, and more negative *emotional responses* than males. In the next section we explore how each of these sources of self-efficacy contribute to students' physics self-efficacy beliefs.

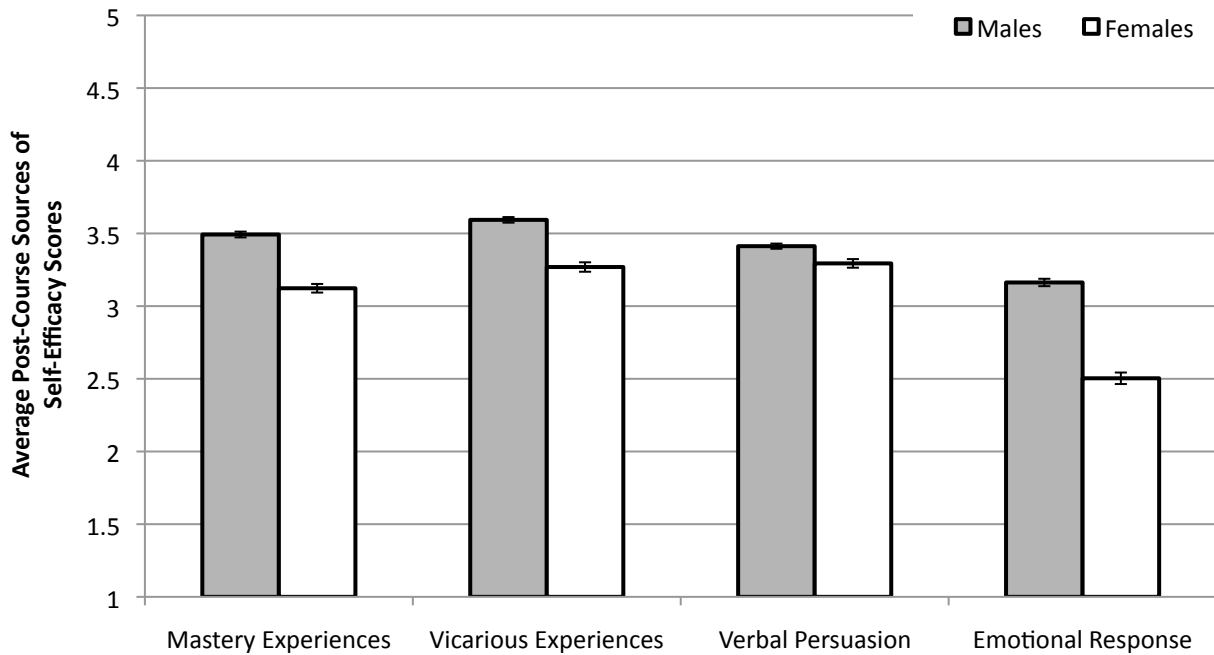


Figure 15. Average male and female scores for each of the four sources of self-efficacy, averaged over all semesters. Error bars represent the standard error on the mean. Females are significantly lower on all sources of self-efficacy.

C. Regression Modeling

Our initial interest in self-efficacy was as a potential predictor of student performance and moderator of performance differences by gender. Having characterized the physics self-efficacy beliefs of males and females, we now ask whether physics self-efficacy beliefs predict student performance in the introductory physics courses. We are also interested in how self-efficacy beliefs develop, particularly what role the four sources of self-efficacy play. To answer these questions, we model student performance or students' self-efficacy beliefs using multiple regression analysis. Details of the multiple regression method can be found in prior chapters. Briefly, we use a combination of independent variables (both categorical and continuous variables) to predict a dependent variable (average exam score, FMCE post-test score, or post-course self-efficacy beliefs).

A large number of independent variables are used in the following analysis, and in order for a student to be included, he/she must have data for each variable. As a consequence, not all students who took the Physics Self-Efficacy and Identity Survey are included in the regression models. Only 63% of the students in the final survey sample were included in the regression sample. As is often the case when non-random sampling procedures are used, students in the regression sample are significantly different (in terms of course grades) than students who are not in the regression sample. Again, this means that we have to take care when generalizing the results beyond the regression sample.

In this section, we report on the regression models that are used to answer the following questions: 1) Does physics self-efficacy predict performance on exams? 2) Does self-efficacy predict performance on the FMCE? and 3) Do the sources of self-efficacy predict self-efficacy? The analysis was completed for students in Physics 1 and Physics 2 separately (since different sets of independent variables are used). We present results here for Physics 1, but similar trends are found in Physics 2. Details of the regression models for Physics 2 are found in Appendix D.

1. Does self-efficacy predict performance on exams?

We are first interested in whether self-efficacy is a useful predictor of exam performance, and particularly if it is useful beyond the academic background factors (like FMCE pre-test score and math score) that we identified previously as being important predictors of student performance. In Table 6 we present four regression models where we have regressed students' average exam score on several predictor variables. Not only do the independent variables in each model significantly predict the dependent variable, but also the change in R^2 for each successive model is statistically significant, meaning that each model accounts for significantly more variance in exam scores than the previous model. In Model 1, the only predictor variable

included is *Female*, a dummy variable with a value of 1 for females and 0 for males. The coefficient of -4.44 in the regression model indicates that the gender difference in exam scores for this sample of students is 4.44%, favoring males. In Model 2, we add covariates that we have previously identified are useful predictors of student performance, as well as variables to account for the semester that students took the course. All of these covariates are significant, meaning that they are all useful predictors of students' exam scores. Also, taking into account these background factors, the difference in exam scores is reduced from -4.44 to -0.11, which is no longer a significant gender difference. This suggests that unlike in prior chapters, where we found that gender gap was only partially accounted for by background, in these semesters, the gender gap in exam scores can be fully accounted for by differences in the mathematics and physics backgrounds of males and females.

In the next stage (Model 3), we enter self-efficacy into the model. Though we included students' pre-course overall physics self-efficacy and each of the four dimensions of physics self-efficacy into the model, only the pre-course *performance* self-efficacy was a significant predictor of students' exams scores. Additionally, students' pre-course *emotional response* score was also a significant predictor of exam scores. Because that was the only source of self-efficacy that we had pre-course data on, it was the only one we could include in the regression models. In the final model (Model 4), we tested all interactions between the independent variables that were included in Model 3 and *Female* to see if any of the independent variables predicted exam scores differently for males and females. The only significant interaction was the *Female* \times *Math Score* interaction, indicating that the slope of the line relating math score and exam score for females was greater than the slope for males. In this final model, which explains 58% of the variance in students' exam scores, the adjusted gender gap in exam scores is not statistically significant.

Table 6. Coefficient estimates and multiple regression model statistics for each regression model. The dependent variable is the average exam score in Physics 1. β is the un-standardized coefficient, β is the standardized coefficient.

	Model 1	Model 2	Model 3	Model 4
Model-Level Statistics				
Multiple R^2	0.020	0.565	0.579	0.584
F statistic p value	<0.001	<0.001	<0.001	<0.001
Std. Error	14.1	9.4	9.3	9.2
Independent Variables				
Constant	B 69.8 β Sig. <0.01	B 64.6 β Sig. <0.01	B 55.1 β Sig. <0.01	B 55.3 β Sig. <0.01
<i>Female</i>	-4.44 -0.14 <0.01	-0.11 0 0.89	0.60 0.02 0.45	0.84 0.03 0.30
FMCE Pretest	0.25 0.47 <0.01	0.24 0.44 <0.01	0.24 0.44 <0.01	0.24 0.44 <0.01
Math Score	5.43 0.36 <0.01	4.99 0.33 <0.01	4.30 0.29 <0.01	4.30 0.29 <0.01
Semester J1-2	-4.69 -0.16 <0.01	-4.38 -0.15 <0.01	-4.40 -0.15 <0.01	-4.40 -0.15 <0.01
Semester K1-2	-11.30 -0.35 <0.01	-10.21 -0.32 <0.01	-10.25 -0.32 <0.01	-10.25 -0.32 <0.01
Pre-Performance SE	1.35 0.07 0.02	1.31 0.07 0.03	1.17 0.08 0.02	1.17 0.08 0.02
Pre-Emotional Response	1.24 0.08 0.01	1.17 0.08 0.01	1.17 0.08 0.01	1.17 0.08 0.01
Female \times Math Score	2.36 0.09 <0.01	2.36 0.09 <0.01	2.36 0.09 <0.01	2.36 0.09 <0.01

2. Does self-efficacy predict performance on the FMCE?

A similar model of student performance emerges when regressing the FMCE post-test on the same independent variables. The five regression models are presented in Table 7. Again, all models are statistically significant, and each model accounts for significantly more variance in FMCE scores than the previous model. According to Model 1, the difference in FMCE post-test scores of males and females in this sample is 11.28%. When background factors are taken into account (Model 2), this gender gap reduces to 2.88%, and is no longer statistically significant [$t(727)=-1.53, p=0.13$]. For this sample of students, about 74% of the gender gap in FMCE post-test scores can be accounted for by prior physics and mathematics understanding.

In Model 3, we include the pre-course *performance* self-efficacy. As in the model of exam scores, pre-course *performance* self-efficacy is a significant predictor of students' FMCE post-test score. In addition, pre-course performance self-efficacy accounts for an additional 12% of the gender gap in post-test scores. At least some of the gender gap in FMCE post-test scores can be attributed to females' lower sense of *performance* self-efficacy.

In Model 4, we find, again, that the *Female* \times *Math Score* interaction is a significant predictor of students' FMCE post-test scores, meaning that the math score predicts FMCE post-test score differently for males and females. This was the only significant interaction.

Table 7. Coefficient estimates and multiple regression model statistics for each regression model. The dependent variable is the FMCE post-test score. B is the un-standardized coefficient, β is the standardized coefficient.

	Model 1	Model 2	Model 3	Model 4	Model 5
Model-Level Statistics					
Multiple R ²	0.031	0.403	0.416	0.422	0.444
F statistic p value	<0.001	<0.001	<0.001	<0.001	<0.001
Std. Error	28.4	22.4	22.2	22.0	21.6
Independent Variables					
Constant	68.2	44.7	26.3	26.7	13.1
	β	β	β	β	β
	<0.01	<0.01	<0.01	<0.01	<0.01
	Sig.	Sig.	Sig.	Sig.	Sig.
<i>Female</i>	-11.28	-2.88	-1.48	-0.92	-0.47
	-0.18	-0.05	-0.02	-0.01	-0.007
	<0.01	0.13	0.44	0.63	0.80
FMCE Pretest		0.53	0.45	0.50	0.43
		0.48	0.45	0.46	0.40
		<0.01	<0.01	<0.01	<0.01
Math Score		6.65	5.95	4.30	3.57
		0.22	0.20	0.14	0.12
		<0.01	<0.01	<0.01	<0.01
Semester J1-2		6.62	6.77	6.75	7.73
		0.11	0.12	0.12	0.13
		<0.01	<0.01	<0.01	<0.01
Semester K1-2		-1.13	0.35	0.31	1.43
		-0.02	0.01	0.01	0.02
		0.60	0.87	0.89	0.51
Pre-Performance SE		4.88	0.13	4.71	2.97
		<0.01	<0.01	<0.01	0.08
Female \times Math Score		5.55	0.10	5.80	0.10
		<0.01	<0.01	<0.01	<0.01
Mastery Experiences		6.56	0.18	6.56	0.18
		<0.01	<0.01	<0.01	<0.01

Because students take the FMCE post-test at the end of the semester (unlike the course exams which are taken throughout the semester), it is reasonable to use students' post-course ratings of the sources of self-efficacy to predict the FMCE post-test score. These ratings are based on students' experiences during the semester, which precede them taking the FMCE post-test. When all four of the sources of self-efficacy scores are included in the model (with all prior independent variables already included) only the *mastery experiences* source is a significant predictor of FMCE post-test scores. The *vicarious experiences* [$\beta=0.07$, $t(721)=1.71$, $p=0.09$], *verbal persuasion* [$\beta=-0.04$, $t(721)=-1.11$, $p=0.27$], and *emotional response* [$\beta=0.03$, $t(721)=0.57$, $p=0.57$] sources are all not significant predictors. Model 5 is the final model, which explains 44% of the variance in students' FMCE post-test scores. Aside from students' prior physics understanding (FMCE pre-test score), the *mastery experience* source is the strongest predictor of FMCE post-test scores ($\beta=0.18$). Students' interpretations of their successes and failures in the physics course, along with their pre-course performance, appear to be important predictors of their end-of-semester conceptual physics performance.

3. Do the sources of self-efficacy predict self-efficacy?

Now that we have seen that students' self-efficacy beliefs, particularly their *performance* self-efficacy beliefs, are important predictors of their performance in the course, we want to know what factors impact students' *performance* self-efficacy. In Table 8 we present four regression models in which post-course *performance* self-efficacy is regressed on a set of independent variables including both student performance and the four sources of self-efficacy. Again, all four regression models are statistically significant, and each model accounts for significantly more variance in post-course *performance* self-efficacy than the previous model. In Model 1, we see that there is a gender difference of 0.49 points (on a 1 to 5 scale) in post-course

performance self-efficacy for this sample of students. In Model 2, students' average exam score is included to account for differences in males' and females' performance in the course. Student performance in the course accounts for 39% of the gender gap in post-course *performance* self-efficacy. Students' actual performance in the course contributes significantly to their post-course *performance* self-efficacy, and also helps account for the gender gap in self-efficacy. In Model 3, students' pre-course *performance* self-efficacy is added to the model. Students' confidence level on tasks related to performing in the course coming into the introductory physics course significantly predicts their confidence at the end of the course, and also accounts for an additional 33% of the gender gap in *performance* self-efficacy.

Table 8. Coefficient estimates and multiple regression model statistics for each regression model. The dependent variable is the post-course *performance self-efficacy* score. *B* is the un-standardized coefficient, β is the standardized coefficient.

	Model 1	Model 2	Model 3	Model 4
Model-Level Statistics				
Multiple R^2	0.051	0.441	0.557	0.732
<i>F</i> statistic <i>p</i> value	<0.001	<0.001	<0.001	<0.001
Std. Error	0.95	0.73	0.65	0.51
Independent Variables				
Constant	<i>B</i> 3.59 <i>β</i> Sig. <0.01	<i>B</i> 0.49 <i>β</i> Sig. <0.01	<i>B</i> -0.77 <i>β</i> Sig. <0.01	<i>B</i> -1.01 <i>β</i> Sig. <0.01
<i>Female</i>	-0.49 -0.23 <0.01	-0.30 -0.12 <0.01	-0.14 -0.07 0.01	-0.01 -0.01 0.76
Average Exam Score	0.04 0.38 <0.01	0.03 0.49 <0.01	0.01 0.49 <0.01	0.19 <0.01
Semester J1-2	0.13 -0.05 0.05	0.09 0.05 0.12	0.06 0.05 0.17	0.03 0.03 0.17
Semester K1-2	0.08 -0.19 0.29	0.11 0.05 0.10	0.04 0.05 0.10	0.02 0.02 0.41
Pre-Performance SE			0.49 0.38 <0.01	0.30 0.24 <0.01
Mastery Experiences				0.24 0.19 <0.01
Vicarious Experiences				0.12 0.10 0.01
Verbal Persuasion				0.12 0.09 <0.01
Emotional Response				0.27 0.29 <0.01

Model 4, which includes all four sources of self-efficacy, is the final model of students' post-course *performance* self-efficacy. The independent variables in this model account for 73% of the variance in *performance* self-efficacy. All four sources are significant predictors of students' post-course *performance* self-efficacy. As Bandura predicted, *mastery experiences* is a strong predictor of self-efficacy ($\beta=0.19$), but the *emotional response* source is the strongest predictor ($\beta=0.29$). More than anything else that is included in the model, students' reported levels of stress, worry, and uneasiness while in the course have the largest impact on their *performance* self-efficacy at the end of the semester. Their pre-course *performance* self-efficacy ($\beta=0.24$), their interpretations of their performance in the course (*mastery experiences*), and students' actual performance in the course ($\beta=0.19$) all have a significant impact on their post-course *performance* self-efficacy. The other two sources of self-efficacy, *vicarious experiences* and *verbal persuasion*, while significant predictors, have a smaller impact on post-course *performance* self-efficacy ($\beta=0.10$ and $\beta=0.09$, respectively). Additionally, in Model 4, the gender difference in students' post-course *performance* self-efficacy is not significant. The set of variables included in the final model completely accounts for the gender gap in *performance* self-efficacy.

There was some evidence in prior literature that the sources of self-efficacy differently impacted self-efficacy for males and females [35,36]. To test whether this was the case in our sample, we included all interactions between *Female* and each of the sources of self-efficacy. None of these interactions was significant [*mastery experiences*: $\beta=-0.05$, $t(714)=-0.07$, $p=0.59$; *vicarious experiences*: $\beta=0.01$, $t(714)=0.01$, $p=0.90$; *verbal persuasion*: $\beta=-0.03$, $t(714)=-0.05$, $p=0.69$; *emotional response*: $\beta=0.06$, $t(714)=0.08$, $p=0.34$], suggesting that there are no differences in how the sources predict self-efficacy for males and females. We do not find

evidence for students in our sample, that any of the sources of self-efficacy are more or less predictive or critical for males or females. We also included measures of students' background understanding of physics and mathematics, but neither was a significant predictor of post-course *performance* self-efficacy once the variables in Model 4 were already included [FMCE pretest: $\beta=0$, $t(716)=-0.13$, $p=0.90$; Math Score: $\beta=-0.002$, $t(716)=-0.06$, $p=0.95$].

4. Summary

Using multiple regression analysis, we were able to show that self-efficacy, specifically *performance* self-efficacy, is an important predictor of student performance in the course, both on exams and on the FMCE post-test, even when already controlling for background factors. Students' pre-course confidence about doing well on exams, completing homework assignments, discussing physics ideas, and answering professor questions significantly contributes to their exam and conceptual survey performance. Additionally, accounting for gender differences in students' pre-course *performance* self-efficacy reduced the gender gap in FMCE post-test scores by an additional 12%, when already taking into account background factors. We also found that students' interpretations of their performance in the course, including their homework scores, how difficult they found exams, and whether they knew the answer to clicker questions, was also an important factor that impacted student performance on the FMCE post-test.

As far as what impacts students' *performance* self-efficacy, we found that all four sources of self-efficacy were significant predictors of *performance* self-efficacy. *Emotional responses* had the largest impact on *performance* self-efficacy, and pre-course *performance* self-efficacy, *mastery experiences*, and students' exam performance also had sizable impacts. Both students' actual course performance and their interpretations of their performance made unique contributions to students' post-course *performance* self-efficacy. Contrary to prior literature, we

did not find any evidence that the critical sources of self-efficacy were different for males and females.

V. Conclusions

In this study, we surveyed students to assess their physics self-efficacy and their interpretations of experiences in the physics course that impact their self-efficacy (sources of self-efficacy). We identified gender differences in students' overall physics self-efficacy coming into the course; females were less confident about completing the tasks necessary for success in the introductory physics courses. However, when we break physics self-efficacy out into four dimensions, pre-course gender differences exist only for the *performance* dimension of physics self-efficacy. While females are less confident than males regarding doing well on exams and in the course, males and females are equally confident about using mathematics in the physics course, getting help from the professor and TAs, and working with other students. The lack of gender differences does not persist through the semester. At the end of the semester, females report less confidence not only in doing well on exams and in the course, but also in using mathematics in physics and getting help from instructors. While self-efficacy beliefs generally decrease over the semester for all students, females have more negative shifts in these dimensions of self-efficacy than males. Additionally, controlling for students' actual exam performance does not fully explain the difference in post-course *performance* self-efficacy between males and females.

Having identified gender differences in students' physics self-efficacy, we also wanted to understand differences in how males and females interpret their experiences in the introductory physics courses that contribute to the development of their self-efficacy beliefs. Given the gender differences in post-course physics self-efficacy beliefs that were observed, we would expect to

find significant gender differences in each of the four sources of self-efficacy. This was the case; we identified significant gender differences on all four sources, though large differences were present only for the *mastery experiences*, *vicarious experiences*, and *emotional response* sources. Males and females were having different experiences and/or interpreting those experiences differently in the introductory physics course, which would lead to differing development of physics self-efficacy beliefs for males and females.

Multiple regression analysis was used to understand what role self-efficacy plays in student performance. *Performance* self-efficacy significantly contributed to students' exam and FMCE performance in the course, even when background factors like prior physics and mathematics understanding are included in the regression model. Also, students' pre-course *emotional responses*, their physiological reactions to prior science and physics courses, were a significant predictor of their exam performance in introductory physics. Of the four sources of self-efficacy, only the *mastery experiences* source was a significant predictor of students' performance on the FMCE post-test.

Regression analysis was also used to understand how the sources of self-efficacy contribute to students' self-efficacy development. All four of the sources of self-efficacy significantly predicted students' post-course *performance* self-efficacy. The *emotional responses* source was the strongest predictor, followed by students' pre-course *performance* self-efficacy, the *mastery experiences* source, and students' exam performance. There was no evidence to suggest that the impact of the sources of self-efficacy differed for males and females.

VI. Discussion

It is clear from this study that males and females interpret their experiences in the introductory physics courses differently, those interpretations result in differences in physics

self-efficacy beliefs, which then contributes to the differences in performance between males and females in the introductory physics courses. The work of other researchers also suggests that self-efficacy is a critical factor affecting students' persistence in science. Taken together, we argue that the self-efficacy beliefs of students need to be addressed in the introductory physics course if we want to increase the performance and retention of females in physics. Being aware of differences between males and females, as well as the negative impact of the introductory courses is the first step.

The next step is to make changes to the structure of the introductory courses in order to allow for more experiences that positively impact students' self-efficacy. Interventions that provide students with opportunities for mastery experiences have been successful at increasing students' self-efficacy beliefs [31,32]. There is also evidence from these studies that reflecting on the mastery experience by justifying why students could be successful in science may be critical in the usefulness of the mastery experiences for increasing self-efficacy. In order to provide mastery experiences for students of all ability levels, exercises in the course need to be aimed at a variety of student background levels, rather than preferentially teaching to the high-performing students. Though the impact of vicarious experiences and verbal persuasion on self-efficacy was smaller than the impact of mastery experiences, it is still important to address these sources of self-efficacy. When students interact with one another they are provided with opportunities to observe their peers being successful. It is important that these interactions are structured in order to facilitate positive interactions that will increase self-efficacy. Creating a classroom community in which the interactions are comfortable, and not threatening, where expression of ideas, even wrong ideas, is appreciated and supported, and where students collaborate with diverse groups of students who have a range of strengths and weaknesses are all ways to provide productive

vicarious experiences for students [40]. Encouragement and verbal support from peers and instructors, while only a small contributor to self-efficacy, is important, and some evidence suggests that verbal persuasion can enhance or undermine the impact of other sources [23].

Gender differences in students' emotional responses to their prior physics courses as well as to the Physics 1 and Physics 2 courses, and the strength of the impact of emotional responses on student performance and students' physics self-efficacy beliefs is somewhat troubling. In one sense, it seems impossible that we, as researchers and teachers, can change students' physiological responses. On the other hand, small changes to the classroom environment and course structure could go a long way towards easing the stress, worry, and uneasiness of students. Supportive, non-competitive environments that stress learning over performing could reduce negative emotional responses among students [55,56,57]. There is some evidence that more interactive and engaging physics courses that focus on developing students' understanding of physics concepts can positively impact females' self-efficacy [40]. Additionally, careful thought should go into how students are assessed in these courses. Given that females have more negative emotional responses to exams, these negative emotional responses contribute to exam performance directly and through self-efficacy beliefs, and that females interpret performance in the course more negatively compared to equally performing males, using exams as a primary way of assessing students should be reconsidered. As the sources of self-efficacy equally impact self-efficacy for males and females, changes to increase the self-efficacy of females will likely be beneficial for all students.

Finally, the large gender differences in students' emotional responses to exams in the course specifically, suggest that identity threat [58,59,60] may be negatively impacting females

introductory physics courses. Identity, identity threat, and a brief intervention to alleviate identity threat among females are the subjects of the next two chapters.

Peer-reviewed publications based on this work

- L. E. Kost, S. J. Pollock and N. D. Finkelstein, “Unpacking Gender Differences in Students’ Perceived Experiences in Introductory Physics” in *Proceedings of the 2009 Physics Education Research Conference*, edited by M. Sabella, C. Henderson and C. Singh (AIP Conference Proceedings, New York, 2009) p. 177.

References (Chapter 4)

- [1] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, *Psychological Review* **84**(2), 191 (1977).
- [2] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice Hall, Englewood Cliffs, NJ, 1986).
- [3] F. Pajares, Self-efficacy beliefs in academic settings, *Review of Educational Research* **66**(4), 543 (1996).
- [4] R. W. Lent and S. D. Brown, On conceptualizing and assessing social cognitive constructs in career research: A measurement guide, *Journal of Career Assessment* **14**, 12 (2006).
- [5] G. Hackett and N. E. Betz, A self-efficacy approach to the career development of women, *Journal of Vocational Behavior* **18**(3), 326 (1981).
- [6] N. E. Betz and G. Hackett, The relationship of career-related self-efficacy expectations to perceived career options in college women and men, *Journal of Counseling Psychology* **28**(5), 399 (1981).
- [7] N. E. Betz and G. Hackett, The relationship of mathematics self-efficacy expectations to the selection of science-based college majors, *Journal of Vocational Behavior* **23**, 329 (1983).
- [8] G. Hackett, The role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis, *Journal of Counseling Psychology* **32**(1), 47 (1985).
- [9] R. W. Lent, H. Sheu, D. Singley, J. A. Schmidt, L. C. Schmidt and C. S. Gloster, Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students, *Journal of Vocational Behavior* **73**, 328 (2008).
- [10] R. W. Lent, S. D. Brown and K. C. Larkin, Self-efficacy in the prediction of academic performance and perceived career options, *Journal of Counseling Psychology* **33**(3), 265 (1986).
- [11] R. W. Lent, S. D. Brown and K. C. Larkin, Comparison of three theoretically derived variables in predicting career and academic behavior: Self-efficacy, interest congruence, and consequence thinking, *Journal of Counseling Psychology* **34**(3), 293 (1987).
- [12] R. W. Lent, F. G. Lopez and K. J. Bieschke, Mathematics self-efficacy: Sources and relation to science-based career choice, *Journal of Counseling Psychology* **38**(4), 424 (1991).
- [13] G. Hackett and N. E. Betz, An exploration of the mathematics self-efficacy/mathematics performance correspondence, *Journal for Research in Mathematics Education* **20**(3), 261 (1989).

- [14] R. W. Lent, S. D. Brown and K. C. Larkin, Relation of self-efficacy expectations to academic achievement and persistence, *Journal of Counseling Psychology* **31**(3), 356 (1984).
- [15] D. Cervone and P. K. Peake, Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior, *Journal of Personality and Social Psychology*, **50**(3), 492 (1986).
- [16] F. Pajares and M. D. Miller, Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis, *Journal of Educational Psychology* **86**(2), 193 (1994).
- [17] F. Pajares and J. Kranzler, Self-efficacy beliefs and general mental ability in mathematical problem-solving, *Contemporary Educational Psychology* **20**, 426 (1995).
- [18] M. M. Chemers, L. Hu and B. F. Garcia, Academic self-efficacy and first-year college student performance and adjustment, *Journal of Educational Psychology* **93**(1), 55 (2001).
- [19] K. D. Multon and S. D. Brown, Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation, *Journal of Counseling Psychology* **38**(1), 30 (1991).
- [20] In this context, effect size is similar to Pearson's correlation coefficient, and can be interpreted as such.
- [21] R. W. Lent, S. D. Brown and G. Hackett, Toward a unifying social cognitive theory of career and academic interest, choice, and performance, *Journal of Vocational Behavior* **45**, 79 (1994).
- [22] P. J. Rottinghaus, L. M. Larson and F. H. Borgen, The relation of self-efficacy and interests: A meta-analysis of 60 samples, *Journal of Vocational Behavior* **62**, 221 (2003).
- [23] E. L. Usher and F. Pajares, Sources of self-efficacy in school: Critical review of the literature and future directions, *Review of Educational Research* **78**(4), 751 (2008).
- [24] A. Bandura, *Self-Efficacy: The Exercise of Control* (Freeman, New York, 1997).
- [25] T. Matsui, K. Matsui and R. Ohnishi, Mechanisms underlying math self-efficacy learning of college students, *Journal of Vocational Behavior* **37**, 225 (1990).
- [26] F. G. Lopez, R. W. Lent, S. D. Brown and P. A. Gore Jr., Role of social-cognitive expectations in high school students' mathematics-related interest and performance, *Journal of Counseling Psychology* **44**(1), 44 (1997).
- [27] S. L. Britner and F. Pajares, Sources of science self-efficacy beliefs of middle school students, *Journal of Research in Science Teaching* **43**(5), 485 (2006).

- [28] M. A. Hutchison, D. K. Follman, M. Sumpter and G. M. Bodner, Factors influencing the self-efficacy beliefs of first-year engineering students, *Journal of Engineering Education*, **95**(1), 39 (2006).
- [29] N. Z. Hampton and E. Mason, Learning disabilities, gender, sources of self-efficacy, self-efficacy beliefs, and academic achievement in high school students, *Journal of School Psychology* **41**, 101 (2003).
- [30] R. W. Lent, F. G. Lopez and K. J. Bieschke, Mathematics self-efficacy: Sources and relation to science-based career choice, *Journal of Counseling Psychology* **38**(4), 424 (1991).
- [31] D. A. Luzzo, P. Hasper, K. A. Albert, M. A. Bibby and E. A. Martinelli Jr., Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students, *Journal of Counseling Psychology* **46**(2), 233 (1999).
- [32] E. D. Cordero, S. H. Porter, T. Israel and M. T. Brown, Math and science pursuits: A self-efficacy intervention comparison study, *Journal of Career Assessment* **18**(4), 362 (2010).
- [33] E. L. Usher and F. Pajares, Sources of academic and self-regulatory efficacy beliefs of entering middle school students, *Contemporary Educational Psychology* **31**, 125 (2006).
- [34] E. L. Usher and F. Pajares, Inviting confidence in school: Invitations as a critical source of the academic self-efficacy beliefs of entering middle school students, *Journal of Invitational Theory and Practice* **12**, 7 (2006).
- [35] A. L. Zeldin and F. Pajares, Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers, *American Educational Research Journal* **37**(1), 215 (2000).
- [36] A. L. Zeldin, S. L. Britner and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, *Journal of Research in Science Teaching* **45**(9), 1036 (2008).
- [37] A. M. Cavallo, W. H. Potter and M. Rozman, Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life-science majors, *School Science and Mathematics* **104**(6), 288 (2004).
- [38] H. Fencl and K. Scheel, Engaging students: An examination of the effects of teaching strategies on self-efficacy and course climate in a nonmajors physics course, *Journal of College Science Teaching* **35**(1), 20 (2005).
- [39] E. Brewster, Modeling theory applied: Modeling Instruction in introductory physics, *American Journal of Physics* **76**(12), 1155 (2008).

- [40] V. Sawtelle, E. Brewe and L. Kramer, “Positive impacts of Modeling Instruction on self-efficacy” in *Proceedings of the 2010 Physics Education Research Conference*, AIP Conference Proceedings Number 1289, edited by C. Singh, M. Sabella and S. Rebello (AIP, New York, 2010), p. 289.
- [41] A. Gungor, A. Eryilmaz and T. Fakioglu, The relationship of freshman’s physics achievement and their related affective characteristics, *Journal of Research in Science Teaching* **44**(8), 1036 (2007).
- [42] K. A. Shaw, “The development of a physics self-efficacy instrument for use in the introductory classroom,” in *Proceedings of the 2003 Physics Education Research Conference*, AIP Conference Proceedings Number 720, edited by J. Marx, S. Franklin and K. Cummings (AIP, New York, 2003), p. 137.
- [43] F. Pajares and M. D. Miller, Mathematics self-efficacy and mathematics performances: The need for specificity of assessment, *Journal of Counseling Psychology* **42**(2), 190 (1995).
- [44] B. J. Zimmerman, Self-efficacy: An essential motive to learn, *Contemporary Educational Psychology* **25**, 82 (2000).
- [45] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, *Physical Review Special Topics – Physics Education Research* **5**, 020103 (2009).
- [46] S. K. Kachigan, *Multivariate Statistical Analysis: A Conceptual Introduction* (Radius Press, New York, 1991).
- [47] J. Stevens, *Applied Multivariate Statistics for the Social Sciences* (Erlbaum, Mahwah, NJ, 2002).
- [48] R. M. Sirkin, *Statistics for the Social Sciences* (SAGE Publications, Thousand Oaks, CA, 1999).
- [49] R. G. Lomax, *An Introduction to Statistical Concepts for Education and Behavioral Sciences* (Lawrence Erlbaum, Mahwah, NJ, 2001).
- [50] W. J. Conover, *Practical Nonparametric Statistics* (John Wiley & Sons, New York, 1999).
- [51] In the J1-2 and K1-2 semesters of Physics 1, we implemented an intervention (discussed in Chapter 6) to increase females’ performance in the course, which could have also impacted students’ self-efficacy. However, we found no significant impact of the intervention on any self-efficacy measures that we collected.
- [52] The grouping of students is done separately for each semester, which means that students that fall into the lowest quintile don’t necessarily have the same average score (since

average exam scores are different each semester), but they all scored in the lowest 20% of their class.

- [53] Summary tables of average values of physics self-efficacy and the sources of self-efficacy can be found in Appendix C.
- [54] The variation in pre-course scores is somewhat surprising, as we would not expect students on average to have very different high school physics experiences. There are several possible explanations: 1) the pre-course questions were phrased differently in the K1-1 and J2-2 semesters compared to the others (asking about prior *science* courses as opposed to prior *physics* courses), 2) there are differences in the fall and spring populations of students which may contribute to how they respond to these questions and 3) the survey is given in the second week of the course, so differences between how professors frame the course in the first week could impact students' responses.
- [55] *Teaching the Majority: Breaking the Gender Barrier in Science, Mathematics, and Engineering*, edited by S. V. Rosser (Teacher's College Press, New York, 1995).
- [56] National Research Council, *To Recruit and Advance: Women Students and Faculty in Science and Engineering* (National Academies Press, Washington, DC, 2006).
- [57] D. Halpern, J. Aronson, N. Reimer, S. Simpkins, J. Star, and K. Wentzel, *Encouraging Girls in Math and Science* NCER 2007-2003. (Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education, 2007). Retrieved from <http://ncer.ed.gov>.
- [58] C. M. Steele, S. J. Spencer and J. Aronson, "Contending with group image: The psychology of stereotype and social identity threat" in *Advances in Experimental Social Psychology* **34**, 379 (2002).
- [59] C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance, *American Psychologist* **52**, 613 (1997).
- [60] S. J. Spencer, C. M. Steele and D. M. Quinn, Stereotype threat and women's math performance, *Journal of Experimental Social Psychology* **35**, 4 (1999).

CHAPTER 5: IDENTITY IN PHYSICS

I. Introduction

In this chapter, we explore students' sense of physics identity, both how they recognize themselves and how they perceive being recognized by others as a physics person. We seek to answer the following research questions: 1) Are there gender differences in students' sense of physics identity? 2) How does students' physics identity change over the course of Physics 1 and Physics 2? and 3) Does identity predict student performance in the course?

We begin by introducing the theory of learning as identity formation, and then outline the prior literature on identity in science. We also introduce the model of physics identity that is used throughout our study, and the prior work that it is based on. We then summarize the research methods that were used in this study, including the development and administration of the Physics Self-Efficacy and Identity Survey, and the statistical methods that were employed. The results are presented in three major sections: students' sense of self-identity, students' recognition by others as a physics person, and regression modeling. We find that males are significantly more likely to self identify and to report that they are recognized by others as a physics person. When asked why they do or do not see themselves as a physics person, most student responses involve either their interest in physics or their abilities in physics. More than half of both males and females report that they felt like less of a physics person at the end of Physics 1 or Physics 2 compared to the beginning of the class. Again, the most common reasons were loss of interest in physics or because they didn't do well in physics. Regression modeling was used to determine whether students' physics identity was a useful predictor of student

performance in the course. Though self-identity was a useful predictor on its own, when measures of students' self-efficacy were included, self-identity was no longer a significant predictor of end-of-semester conceptual survey performance. Implications for the introductory physics courses and how they can be changed to positively impact students' development of physics identity are discussed.

II. Background

A. Learning as Identity Formation

Traditionally, learning is seen as the acquisition of knowledge and the application of that knowledge to situations unlike the one in which it was learned. More recently, researchers have viewed learning as increasing participation in a community of practice, where a community of practice is a group of people all working together towards an agreed-upon goal, using tools, resources, and ways of doing things that they have appropriated or developed [1,2]. By participating in activities that are increasingly valued by and essential to the community of practice, people learn how to be a member of that community of practice. This view of learning redirects the focus from *what* students are learning, to *who* students are learning to be, and brings identity to the fore. As Wenger describes, "Because learning transforms who we are and what we can do, it is an experience of identity. It is not just an accumulation of skills and information, but a process of becoming—to become a certain person or, conversely, to avoid becoming a certain person," [2, p. 215]. From this perspective, learning is the process of identity formation. People develop the knowledge, competence, ways of talking, ways of acting, ways of using tools, and ways of being that are required for participation in a community of practice [2,3].

Many studies have demonstrated a link between students' identities and their engagement and learning in science or math [4,5,6,7,8,9,10,11]. Boaler and Greeno [10,12] studied two types

of AP Calculus classes, traditional courses and discussion-oriented courses, where students engaged in discussion about mathematics and had to work together, generate questions and ideas, and judge the validity of ideas. Though students in the two types of classes achieved at similar levels, students in the traditional classes reported that they didn't like mathematics and they didn't want to study it further. The traditional course positioned students solely as receivers of knowledge, and this identity conflicted with the students' developing identities as independent thinkers who generated their own ideas. In contrast, in the discussion-oriented class, students were required to come up with their own ideas and create their own understandings, a type of participation that was more in line with their identity. Brickhouse, Lowery and Schultz [13] studied how four middle-school girls engaged in their science classes. Though all four girls expressed an interest in science and believed that they were good at science, their identities and views of who they were resulted in a variety of ways, some productive and some not, of engaging in the science class. They found that the girls who took on the socially accepted "good student" identity (following teacher's instructions, completing required tasks, being attentive, engaging with the material) were able to participate successfully in school science, but those girls who violated the gender-norms of a "good student" by being disruptive and opinionated, following her own interests and being disengaged with the class material, or demanding to be heard and seen as a leader in the class, were less likely to be seen as good at school science.

Brickhouse [14] argues that in order to understand the lack of women in science and to address the disparity, "we need to know how students engage in science and how this is related to who they are and who they want to be," [p. 286]. It is not enough to examine whether males and females are learning the knowledge necessary to succeed in physics, we have to understand how they view themselves, how they view physicists, and whether the two are compatible. Prior

research has shown that the identities of scientists that are conveyed in school are often not compatible with female students' identities [3,10,13,15,16,17,18,19]. Carlone [15] found that girls in an Active Physics high school class rejected the meaning of scientist that was promoted by the curriculum, because it conflicted with their desire to maintain a "good student" identity. Many of the girls in the class were taking physics to earn a good grade, which would look good on their college applications. Going into the class, they expected to take on the traditional role of a "good student" that had been successful for them in the past. The curriculum, however, required students to understand science concepts in a meaningful way, be active and hands-on in laboratory activities, and to work with other students. Carlone argued that, "a transformation from 'good student' as listener, memorize, and recipient of knowledge to 'good student' as active, hard-worker, problem-solver, troubleshooter, and producer of knowledge is difficult," [p. 404]. The new "good student" identity was seen by the girls in the class as more demanding, and their perceived reward (a good grade) was the same as in a traditional course.

Researchers have also studied the relationship between identity and persistence in science, and found the two to be correlated [20,21,22,23]. Carlone and Johnson [20] found that women of color in their study tended to develop one of three science identities: research scientist (those women who studied science to learn more about the world), altruistic scientist (women who redefined and studied science as a way of satisfying their altruistic values), or disrupted scientist (women who at one time or another felt overlooked or neglected by others within science). Though all the women in their study were successful in science, women with an altruistic scientist identity had to redefine what it meant to do science in order to be successful, and women with a disrupted scientist identity faced more obstacles along their path to becoming a scientist and over time felt more disconnected with science. Hazari *et al.* [21] found that

physics identity correlated strongly with students' ratings of how likely they were of choosing a physical sciences career. The work outlined in this section suggests that understanding identity and the identity development of students in the introductory physics course is critical for understanding how they engage in the course and whether they choose to pursue physics.

B. *Model of Physics Identity*

For our study, we draw from the work of Carlone and Johnson [20], who developed a model of science identity to understand the experiences of successful women of color in their science-related schooling and careers. Their model of science identity was made up of three components: competence (knowledge and understanding of science), performance (social performances of relevant scientific practices), and recognition (recognizing oneself and getting recognized by others as a “science person”). Hazari, *et al.* [21] later built upon this model by making explicit a fourth component, interest (desire/curiosity to think about and understand science), which Carlone and Johnson had included as part of recognition by oneself. Carlone and Johnson claim that someone with a strong science identity would rate him/herself highly and would be rated highly by others in all of these components. It is possible, however, to have “various degrees and different configurations of science identity,” [p. 1190]. For instance, some students may be competent and excellent performers, but they may not be recognized by others as someone that does science. Carlone and Johnson also make the assumption that a person's racial, ethnic, and gender identities interact with their science identity. For women in their study, Carlone and Johnson found that the women's racial, ethnic, and gender identities interacted most critically with the recognition component of science identity, specifically recognition by others. They claim it is easiest to get recognized as a science person if you look, act, and talk like current members of the science discipline. The model of physics identity that we use in our

study, borrowing from both Carlone and Johnson [20] and Hazari *et al.* [21], is presented in Figure 1.

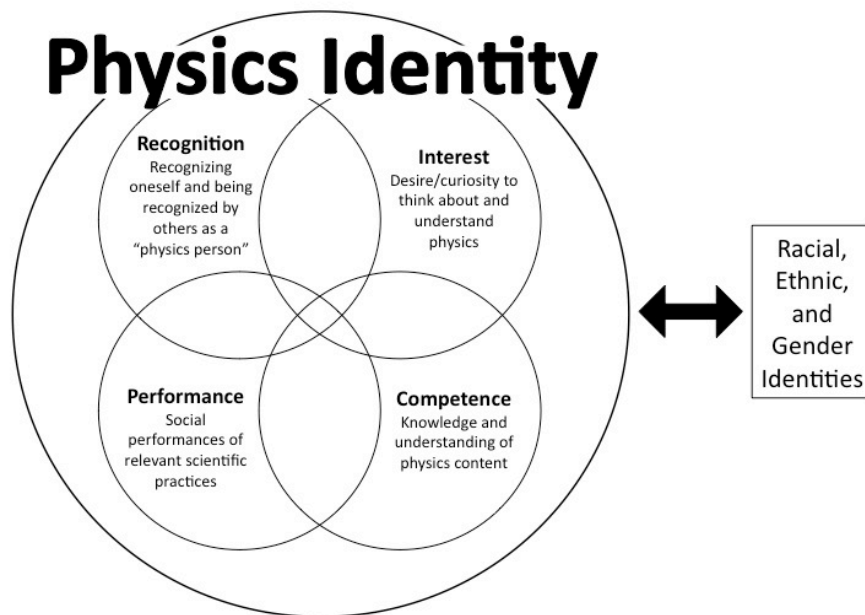


Figure 1. Model of physics identity that guided our investigation of students' physics identity.

The distinction between competence and performance in this model is somewhat unclear. Competence is intended to get at what students know and understand about physics and is measured, according to Carlone and Johnson [20], by test scores and grades. However, we argue, that taking tests and earning grades is a kind of performance. In that way, it can be difficult (or maybe impossible) to truly distinguish what people know from what they show that they know (on tests, homework, or in conversations). For our purposes in this chapter, it is not important to distinguish between competence and performance, and we will consider them to be deeply intertwined, if not one in the same. That being the case, in Chapters 2 and 3 of this thesis, the competence/performance component of physics identity was examined by looking at the exam scores, course grades, and conceptual survey scores of males and females. Both Carlone and

Johnson and Hazari, *et al.* put forth that it is both students' perceptions or beliefs about each component and other people's perceptions about each component that are relevant for determining an individual's identity. In Chapter 4 of this thesis, students' beliefs about their ability to be competent in physics and to perform relevant physics tasks were examined. Having already explored the competence and performance components of identity, both in how others perceive students' abilities (through exam scores and course grades) and how students' perceive their own abilities (through self-efficacy), we now focus on the recognition component of physics identity in this study. Carlone and Johnson [20], as well as others [11,21,24], have identified recognition as a critical component of science identity. In their study, Carlone and Johnson found that receiving recognition from family, teachers, and peers separated those who developed strong science identities from those who did not. In this chapter, we explore the recognition component of the physics identities of students in the introductory physics courses.

III. Research Methods

The goal of this study is to determine to what extent males and females identify and feel recognized by others as a physics person in the introductory physics courses. We also want to understand how identity impacts student performance. Guided by the literature outlined in the previous section, a survey was developed by the author to probe students' sense of physics identity. In this section, we discuss the items from the survey that were used to assess students' physics identity. The items that we use in this analysis are from the same survey that was used to probe students' self-efficacy (in the previous chapter). The survey administration, sample of students, and statistical methodology are the same as in Chapter 4.

The Physics Self-Efficacy and Identity Survey was designed to probe students' sense of physics self-efficacy, the sources of self-efficacy (both discussed in Chapter 4), and students'

sense of physics identity. These three areas comprised the three major sections of the survey. Here, we will discuss the design of the physics identity section of the survey.

Six multiple-choice items were generated by the author to assess students' recognition by themselves and others as a "physics person". The phrase "physics person" has been used in prior research to assess students' sense of identity [21]. The items related to identity on the Physics Self-Efficacy and Identity Survey included the following statements:

- 1) I see myself as a physics person.
- 2) I feel like I could be a good physicist.
- 3) Other students recognize me as a physics person.
- 4) A teacher encouraged me to pursue physics.
- 5) A teacher has told me that I'm good at physics.
- 6) My parents have encouraged me to pursue physics.

Students agreed or disagreed with each statement on a Likert-like scale where the options were: Strongly Disagree, Somewhat Disagree, Neutral, Somewhat Agree, and Strongly Agree. These items were asked on both the pre- and post-course surveys. By asking these questions on both the pre- and post-surveys, shifts in physics identity over the course of the physics class could be examined. Because of concerns about the length of the survey, only four of the six items (#1, #2, #3, and #5) were included in the survey given in the K1-2 and G2-4 semesters. These four items were chosen based on preliminary analysis of which items showed gender differences.

In addition to these multiple-choice items, there were several open-response questions included on the surveys. On the pre-course survey, students were asked:

- Do you see yourself as a science person? Why or why not?
- Do you see yourself as a physics person? Why or why not?

- What are you planning to major in? What made you decide to choose that major? Are there any specific experiences that you recall that helped you decide?
- Did you ever consider being a physics major? What made you decide to choose or not choose physics?

On the post-course survey, students were asked:

- Do you see yourself as a physics person? Why or why not?
- Has your answer to the above question changed since the beginning of the semester? Why or why not?
- Do you feel like you could be a good physicist? Why or why not?

All of these open-response questions were asked only on the surveys given in the K1-1, J1-2, J2-2, and N2-2 semesters. Student responses to some of these questions in the K1-1 semester of Physics 1 were analyzed for this study.

IV. Results

When looking at the six items related to the recognition component of physics identity, we break the items into two groups: self-identity and recognition by others. We discuss results from each source of recognition in this section. Throughout this section, when analyzing student responses to individual survey items, statistical comparisons were made using all five answer options that students were presented with. For simplicity in representing distributions of student responses, the Strongly Disagree and the Somewhat Disagree answer options were collapsed into Disagree, and the Strongly Agree and Somewhat Agree answer options were collapsed into Agree in all plots of responses to individual items. Additionally, student responses from all three semesters of each course were combined.

A. Students' Sense of Self-Identity

There were two items related to students' own recognition of themselves as a physics person. We start by comparing the distribution of male and female responses to each of these items on the pre- and post-course surveys in Physics 1 and Physics 2.

1. I see myself as a physics person.

The distribution of responses to the statement *I see myself as a physics person* is presented in Figure 2. Males are significantly more likely to agree with the statement *I see myself as a physics person* than females in both Physics 1 and Physics 2, at both the beginning and end of the semester [all p 's<0.01]. At the end of Physics 1, about 50% of the female students disagree with the statement *I see myself as a physics person*, and that level of disagreement persists into and throughout Physics 2. Conversely, only between 20% and 30% of males disagree with the statement.

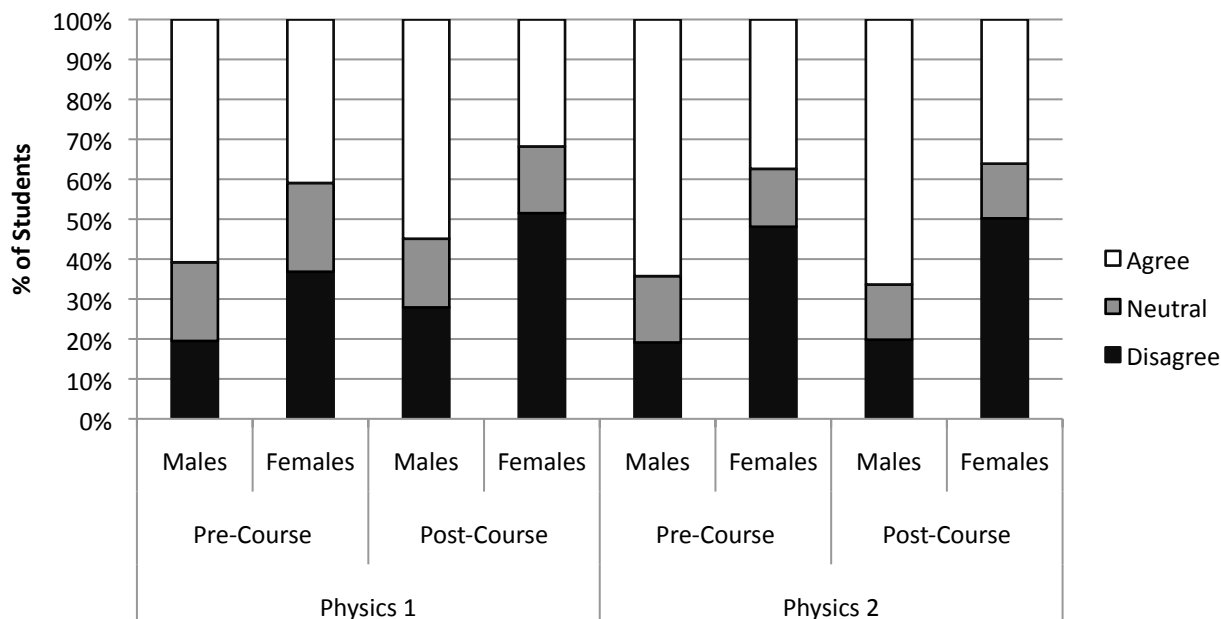


Figure 2. Distribution of male and female responses on the pre- and post-course surveys in Physics 1 and Physics 2 to the statement: *I see myself as a physics person*. In both Physics 1 and Physics 2, on both the pre- and post-course surveys, males are significantly more likely than females ($p < 0.01$) to agree with the statement. Additionally, in Physics 1, both males and females shift towards disagreeing with the statement ($p < 0.01$) from pre- to post-course, but in Physics 2, there is no significant shift ($p > 0.1$) in the distributions of responses.

We also compare the distribution of responses by males and females on the pre- and post-course surveys to see if their responses changed over the semester. In Physics 1, both males and females shifted significantly towards disagreeing with the statement [males: $Z=-4.37$, $p<0.01$; females: $Z=-5.16$, $p<0.01$ [25]]. However in Physics 2, there were no significant differences between the distributions of responses at the beginning and end of the semester [males: $Z=-1.47$, $p=0.14$; females: $Z=-1.43$, $p=0.15$]. While in Physics 1, students report that they see themselves less as a physics person over the course of the semester, there is no change in Physics 2.

In addition to the multiple-choice data collected on this item, we also have students' open responses to the question *Do you see yourself as a physics person? Why or why not?* The pre-course data from the K1-1 semester of Physics 1 was analyzed. Students' pre-course responses to this question were first coded into four categories: yes, no, kind of, and I don't know. Some students answered both yes and no, and so their responses were coded in both categories. Males and females differed significantly in their distribution of responses among the four categories [$\chi^2(3, N=414)=26.8$, $p<0.01$]. In line with students' multiple-choice responses, 67% of males versus 43% of females answered yes, that they did see themselves as a physics person. 24% of males and 40% of females answered no, that they did not see themselves as a physics person. Only about 5% of both males and females said that they sort of saw themselves as a physics person. The remaining students, 6% of males and 17% of females, said that they didn't know if they saw themselves as a physics person, most often because they hadn't taken a physics course before or didn't really have a good idea of what physics was.

The second step in coding students' responses to whether they saw themselves as a physics person, was to categorize their reasons for either being or not being a physics person. There were four main reasons that students gave for seeing themselves as a physics person. Some

students' responses fell into more than one category, and could be double-coded. The distribution of males and females across the four reasons did not differ significantly [$\chi^2(3, N=267)=0.90, p=0.82$]. The most popular reason students gave for seeing themselves as a physics person was that they wanted to know how the world around them worked (54% of females and 43% of males). Students gave responses similar to,

Yes, I see myself as a physics person because understanding physics will help me understand more about the world and my surroundings.

Yes. Physics is very practical and is uses [sic] in everything that we use in everyday life. We are always surrounded by physics whether we like it or not. Physics is very easy to apply to everything.

Another popular reason given was that students liked, enjoyed, or were interested in physics. 44% of both males and females gave this reason for being a physics person. Students who gave this reason wrote,

I do see myself as a physics person in the sense that I enjoy physics and like doing physics problems.

I do see myself as a physics person because it's interesting to me to understand about the movement of objects such as how much work is needed to move a block from point a to point b. There are so many real-life applications.

The other two reasons that students gave for seeing themselves as a physics person were that they were good at physics (17% of females and 10% of males) or that they liked or were good at mathematics (19% of females and 16% of males).

In examining the reasons students gave for why they did not see themselves as a physics person, we identified five main reasons. The distribution of males and females across the five reasons did not differ significantly [$\chi^2(4, N=103)=1.18, p=0.95$]. The most popular reason given

was that students said they were not a physics person because they were not good at physics (31% of females and 28% of males). Another popular reason was that students were more interested in something else (24% of females and 18% of males), giving responses such as,

No. I've always been more interested in the life sciences and in the humanities.

About 20% of both males and females said they were not a physics person because they didn't like, enjoy, or were not interested in physics. Other reasons given for not being a physics person were because they were not good at or didn't like mathematics (12% of females and 10% of males) or they didn't know what physics was (12% of females and 16% of males).

Students gave similar reasons for being and not being a physics person at the end of the course, on the post-course survey. At the end of the semester students were also asked whether their answer to the question of whether they saw themselves as a physics person had changed. In examining the responses from students who reported that their answer had changed, 57% of males and 65% of females reported that they saw themselves as less of a physics person at the end of the course. The difference between males and females is not significant [$\chi^2(1, N=146)=0.85, p=0.36$]. Most students said they were less interested in physics or they realized that they weren't good at physics after taking Physics 1. Some example responses are,

I have lost interest in physics. This physics class isnt fun, its to technical and formal. Physics is fun when you know what youre doing, just knowing the formulas doesnt do anything, which is why I dislike this class; it stresses the formulas way too much. [sic]

Physics has become more difficult as the semester has gone on and I feel that some of the lessons don't delve into real life examples enough for me to conceptualize the problem. As a result, I question my ability to be a physics person because the subject has become more difficult.

43% of males and 35% of females who reported that their answer had changed, said that they felt like more of a physics person. Again, most of these students said they felt like more of a physics person because they were more interested in physics, because they had learned more physics, or because they felt more confident in physics. Some example responses of students who said they became more of a physics person over the course of Physics 1 are,

This semester, I learned about physics a lot more than I did in high school. I know what the physics concepts are saying and how they apply to solving physics problems.

I didn't see myself as a "physics person" at the beginning of the semester but I knew I was capable of doing well in physics. Now I feel more of a connection to the subject and its role in everyday life and my future career.

I am more confident in myself in physics, at least in certain areas like kinematics. Other areas I feel that I have not changed. However in general, I feel like that I am more confident.

From these open-response questions, it seems that while female students are less likely to see themselves as a physics person, males and females give the same reasons for why they are and are not physics people. Students regard themselves as a physics person based primarily on whether they are interested in physics and/or whether they are good at physics.

2. I feel like I could be a good physicist.

The second item related to students' recognition of themselves as a person that does physics was *I feel like I could be a good physicist*. Student responses to this item are presented in Figure 3. Again, as with the previous statement, in both Physics 1 and Physics 2, on both the pre- and post-course surveys, males were more likely to agree with this statement than females [all p 's < 0.01]. At the end of Physics 1 and Physics 2, only about 40% of females agreed with this statement, compared to 60% to 70% of males.

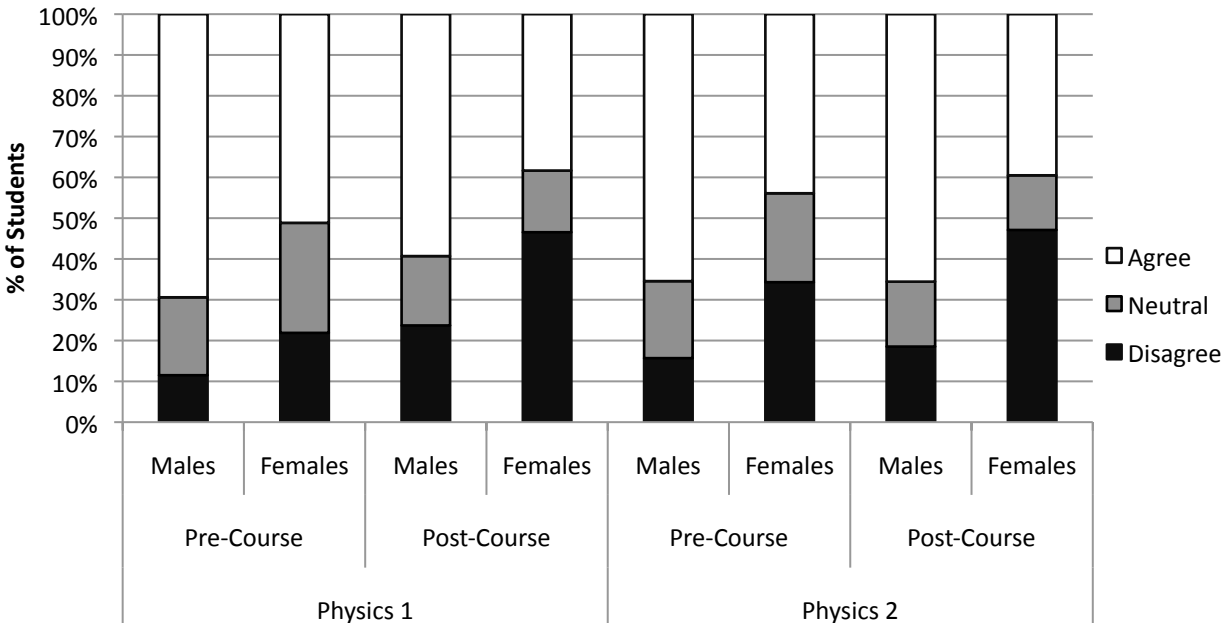


Figure 3. Distribution of male and female responses on the pre- and post-course surveys in Physics 1 and Physics 2 to the statement: *I feel like I could be a good physicist*. The five answer options were collapsed to three categories. In both Physics 1 and Physics 2, on both the pre- and post-course surveys, males are significantly more likely than females ($p < 0.01$) to agree with the statement. Additionally, in Physics 1, both males and females shift towards disagreeing with the statement ($p < 0.01$) from pre- to post-course, but in Physics 2, there is no significant shift ($p > 0.1$) for males in the distributions of responses, but there is for females ($p < 0.01$).

We also compared the pre- to the post-course distributions of responses for males and females. In Physics 1, both males and females shifted towards disagreeing with the statement *I feel like I could be a good physicist* [males: $Z = -8.30$, $p < 0.01$; females: $Z = -6.30$, $p < 0.01$]. In Physics 2, males' responses did not change from pre- to post-course, but females' responses shifted significantly towards disagreement [males: $Z = -1.42$, $p = 0.16$; females: $Z = -3.29$, $p < 0.01$]. Over the course of Physics 1, and for females in Physics 2, students became less likely to agree that they could be a good physicist.

On the post-course survey, students were asked the open-response question *Do you feel like you could be a good physicist? Why or why not?* Students responses were first categorized as either yes, no, or maybe. The response patterns of males and females were significantly different [$\chi^2(2, N=487) = 9.16$, $p = 0.01$ and when only comparing yes/no responses, $\chi^2(1, N=446) = 8.92$,

$p < 0.01$]. 57% of males compared to 43% of females reported that they could be a good physicist. 35% of males and 50% of females reported that they would not be a good physicist. The reasons students gave for feeling like they could or could not be a good physicist were similar to the reasons they gave for seeing themselves or not seeing themselves as a physics person. The most popular responses were that students were or were not interested in physics or were or were not good at physics. Again, there were no significant gender differences in the reasons that students gave for not being a good physicist [$\chi^2(4, N=177)=6.04, p=0.20$]. Some other common reasons that students gave for not feeling like they could be a good physicist, were that they weren't good enough at mathematics, they didn't do well in the course, or that it would be too hard or require too much work. Some students (13% of males and 22% of females) said that they didn't have specific skills that are required for physics, like paying attention to details and being patient. Some examples are,

I feel I could not become a good physicist because I often don't have the patience to solve some of the more complex problems posed in the physics classes I have taken.

I do not feel like I would be a good physicist, because I would forget to take into consideration all of the different concepts that could act on one situation at any time.

I don't think I would. I don't see myself having the capability to think and be innovative in my research.

Not at all. My mind just doesn't seem to work in a physicist's way. My brain can't comprehend all of the material as quickly as a physicist's would.

Other students (12% of males and 8% of females) gave specific aspects of physics that didn't appeal to them, like there was too much theory or too many calculations. For instance,

i prefer to be a problem solver rather than a postulator/experimentor. [sic]

I don't feel like I have enough of a knack to make physics my profession; [...] as for doing research in the field and discovering "new" physics, I'm not certain that I would have enough of a competitive edge.

I do not think I would be a good physicist. That is what my father does. I could not handle the stress and pressure of the job.

Not really, there is too much theory, I am more of a hands-on person.

No, I do problem solving, not theory.

Though I like learning the subject matter, I do not see me as a physics person, because I would much rather be involved in a profession without intense calculations, and more people-to-people interaction.

In many cases, students' comments suggest they have a view of physics that might not align with experts' views. These comments could be a response to how the course was taught in this particular semester. Analysis of student responses in other semesters is needed to see if these trends are widespread. Again, as with the prior open-response questions, while we see gender differences in students' multiple-choice responses, males and females give the same reasons for their responses.

3. Average Self-Identity

Student responses are averaged across the two multiple-choice items to get a measure of students' self-identity, or how much they recognize themselves as a physics person [26]. Pre-course average self-identity scores of males and females are presented in Figure 4. In all semesters, females have a significantly lower sense of physics self-identity than males [all p 's < 0.01]. At the beginning of the Physics 1 and Physics 2 course, females are less likely than males to see themselves as someone that can do physics.

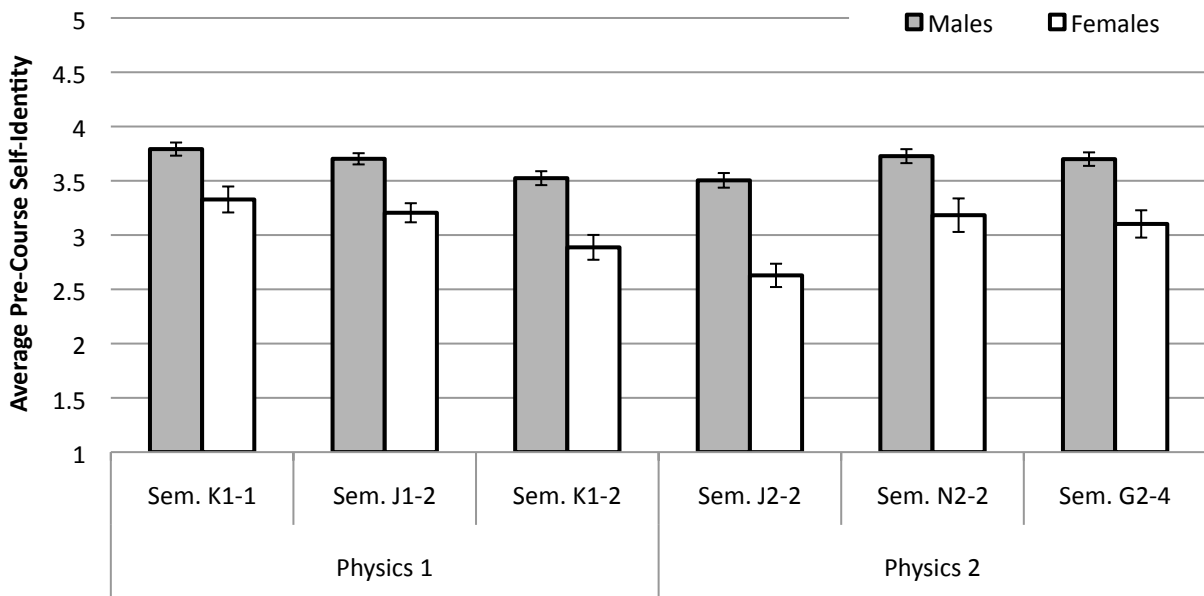


Figure 4. Male and female average pre-course self-identity score for each semester of Physics 1 and Physics 2. In all semesters, males have significantly higher physics self-identity than females at the beginning of the semester (all $p < 0.01$).

Figure 5 presents the effect size of shifts in students' physics self-identity over the course of Physics 1 or Physics 2. In Physics 1, both male and female students shift significantly towards a lower sense of physics self-identity [all p 's < 0.01]. For males, the effect size of the shifts is around -0.3 (a small effect size), while for females the effect size of the shifts is between about -0.3 and -0.6 (small to medium effect sizes). In the K1-1 and J1-2 semesters, there is no significant gender difference in the size of the negative shifts [$t(326) = 0.68$, $p = 0.50$ and $t(475) = 1.41$, $p = 0.16$, respectively], but in the K1-2 semester, females have negative shifts that are more than twice that of males [$t(300) = 2.55$, $p = 0.01$]. In Physics 2, the shifts are much smaller and not always statistically significant. The effect sizes of shifts for males are between about -0.1 and 0.1, while the effect sizes of shifts for females are between about -0.1 and -0.2 (all are small effect sizes). While we observe small to medium negative effect size shifts in Physics 1, the shifts in Physics 2 are small or insignificant.

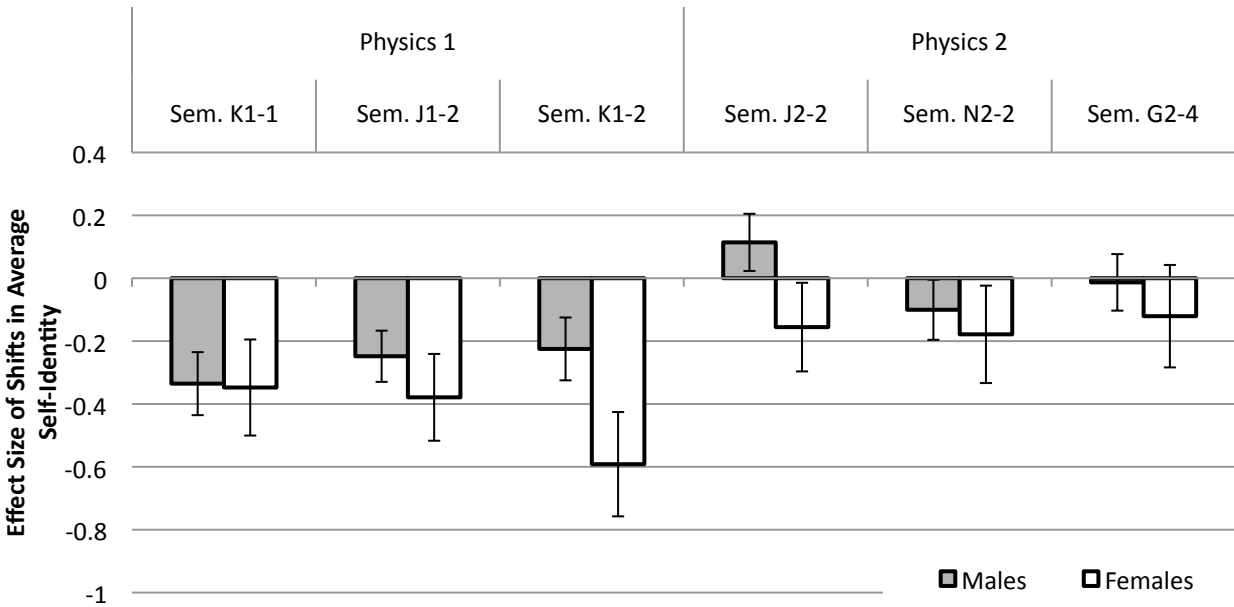


Figure 5. Effect size of shifts in students' self-identity over the course of Physics 1 or Physics 2. The effect size of the shifts is given by $ES = (post - pre) / SD_{pre}$. Error bars are given by $(SE_{pre}^2 + SE_{post}^2)^{1/2} / SD_{pre}$. In all three semesters of Physics 1 both males and females significantly shift towards a lower sense of physics self-identity (all $p < 0.01$). In Physics 2, males do not shift significantly or shift towards a higher sense of physics self-identity (Fall 2009). Females shift significantly toward a lower sense of physics self-identity in two of the three semesters.

We know from the analysis of distributions of students' declared major by gender that males and females have different majors in Physics 1 and Physics 2. Females are less likely to be physics and engineering majors and more likely to be other science majors. The difference in majors between males and females could be responsible for the gender gaps in physics self-identity that we observe. Figure 6 presents the average physics self-identity of males and females in Physics 1 and Physics 2 broken out by students' declared major. This allows a comparison of the physics self-identities of students with the same major. First, for both males and females, there are significant differences in students' pre-course self-identity by major [via ANOVA, both $p < 0.01$]. As we would expect, male and female physics majors had significantly higher physics self-identity than did males and females from other majors. In comparing males to females within each major, in almost all cases, males have significantly higher physics self-identity than females. The exceptions are physics majors in both Physics 1 and Physics 2, non-science majors

in Physics 1, and undeclared majors in Physics 2, who all had no significant gender differences in physics self-identity. While accounting for student major does account for some of the gender difference in students' pre-course physics self-identity, it does not fully explain why females report lower physics self-identity than males.

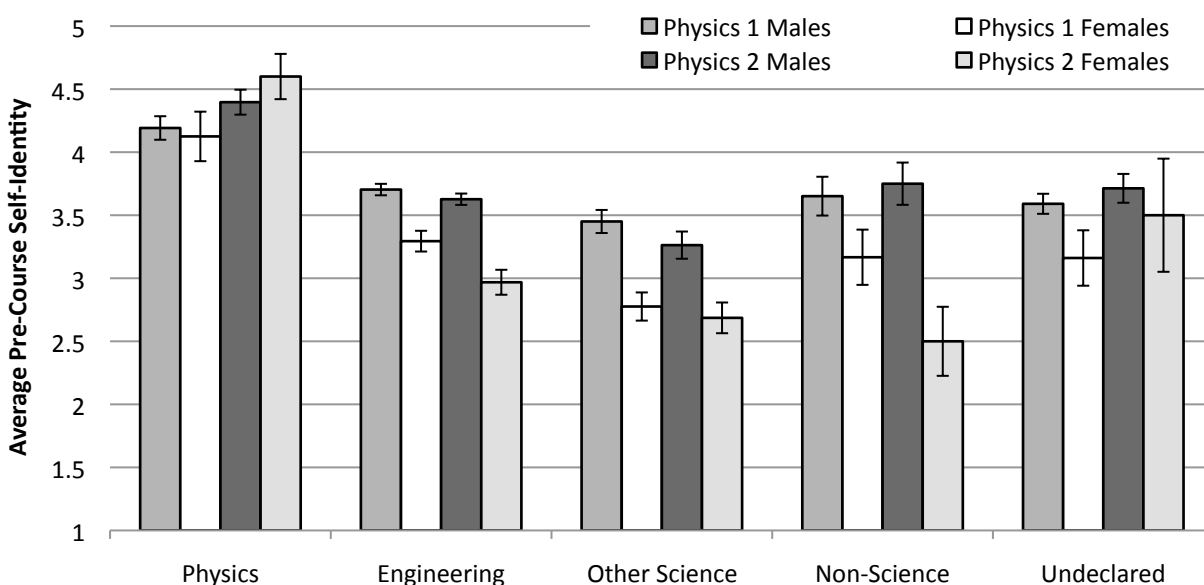


Figure 6. Average pre-course physics self-identity of males and females in Physics 1 and Physics 2 by declared major. For all majors except physics, females have a significantly lower sense of physics self-identity than males.

We also explored how students' self-identity shifted over the semester for different student majors (see Figure 7). In Physics 1, both males and females of all student majors shift toward a lower sense of physics self-identity. These negative shifts have an effect size between -0.2 and -0.6 (small to medium effect sizes). In Physics 2, however, the shifts are much smaller for all students (between about 0.05 and -0.2), except for female physics majors whose negative shifts have an effect size of -0.6. The larger negative shifts of female physics majors over the semester result in females having a lower sense of physics self-identity than males at the end of the semester (the post-course gender differences are not statistically significant though, given the small number of female physics majors, 16 in Physics 1 and 10 in Physics 2). It appears that for

female physics majors, those female students who come into the introductory physics courses wanting to study physics, their sense of being a physics person is negatively impacted in the introductory physics courses.

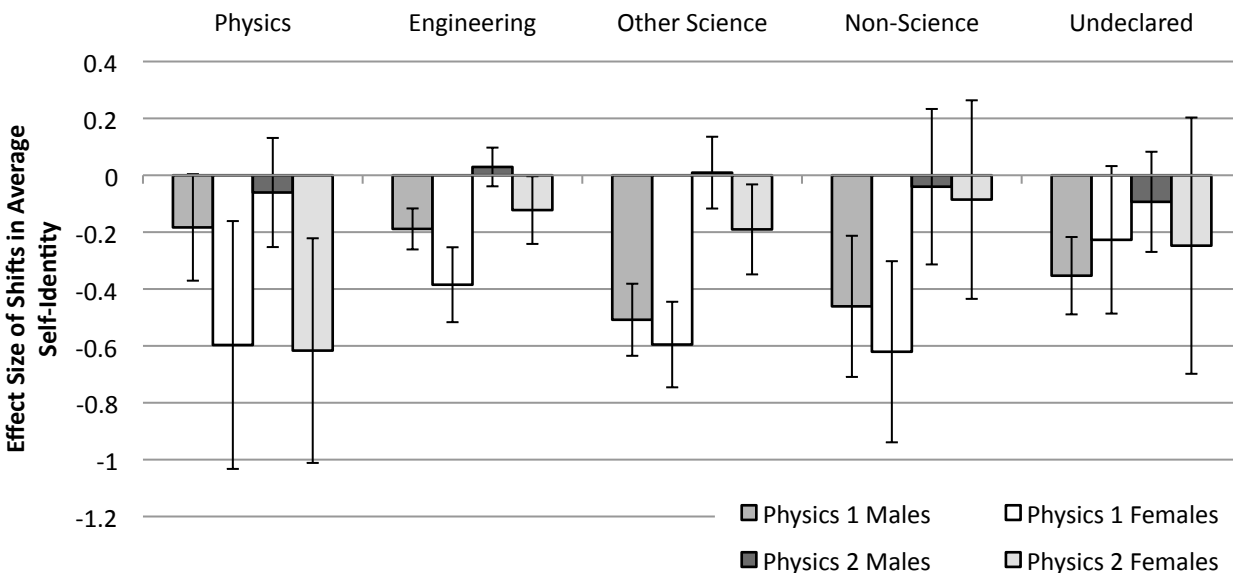


Figure 7. Effect size of shifts in student's physics self-identity by major in Physics 1 and Physics 2. The effect size of the shifts is given by $ES = (post - pre) / SD_{pre}$. Error bars are given by $(SE_{pre}^2 + SE_{post}^2)^{1/2} / SD_{pre}$.

B. Students' Sense of Recognition By Others

Four survey items were related to students' perceptions of being recognized by others, teachers, peers, and parents, as someone that can do physics. We first examine students' responses to each individual item, and then look at males' and females' average recognition by others.

1. Other students recognize me as a physics person.

Students were asked to agree or disagree with the statement, *Other students recognize me as a physics person*. As presented in Figure 8, females are significantly less likely to agree with this statement in both Physics 1 and Physics 2, at both the beginning and end of the semester [all p 's < 0.01]. In Physics 1, only about 30% of females agree that other students recognize them as a

physics person, while about 50% of males agree with the statement. In Physics 2, about 35% of females agree with the statement, while between 55% and 65% of males agree with the statement. In terms of being recognized by peers as someone that is good at physics, only about $\frac{1}{3}$ of the females students, compared to $\frac{1}{2}$ of the male students, report that other students see them as a physics person.

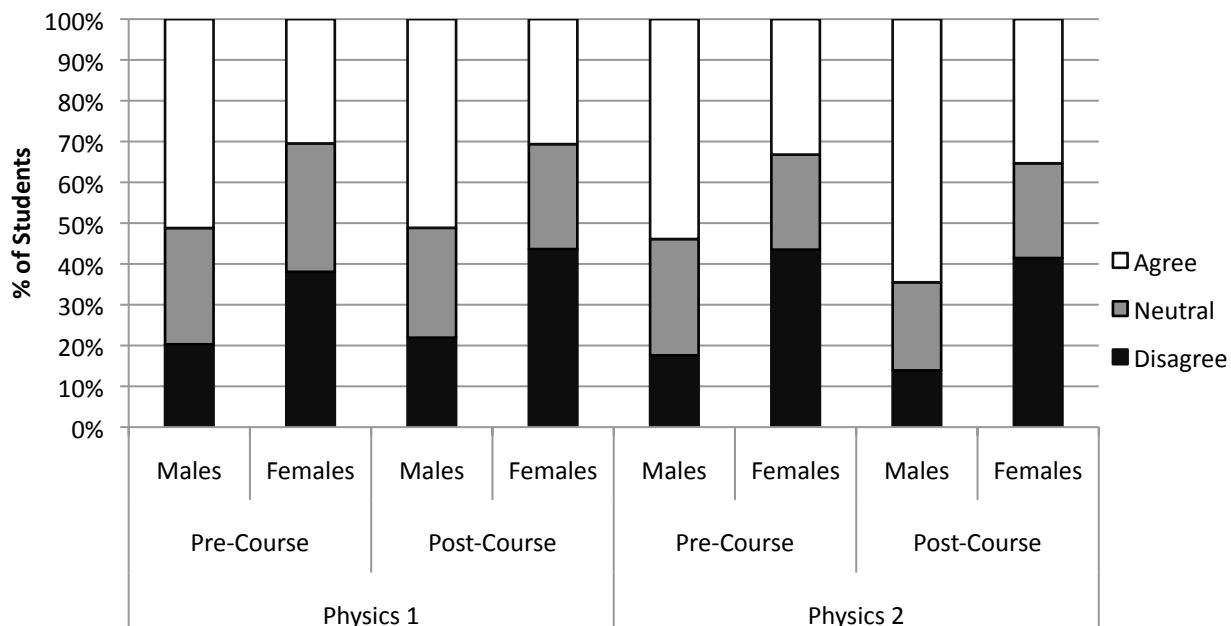


Figure 8. Distribution of male and female responses on the pre- and post-course surveys in Physics 1 and Physics 2 to the statement: *Other students recognize me as a physics person*. The five answer options were collapsed to three categories. In both Physics 1 and Physics 2, on both the pre- and post-course surveys, males are significantly more likely than females ($p < 0.01$) to agree with the statement. Except for males in Physics 2, the distributions of students' responses to this item do not change significantly from pre- to post-course.

There are very few shifts in students' responses to this question over the course of the semester. In Physics 1, neither males nor females have significantly different distributions of responses between the pre- and post-course surveys [males: $Z = -0.36$, $p = 0.72$; females: $Z = -1.04$, $p = 0.30$]. In Physics 2, the female distribution of responses to the statement *Other students recognize me as a physics person* does not change significantly [females: $Z = -1.16$, $p = 0.24$], but the male distribution does [males: $Z = -5.59$, $p < 0.01$]. At the end of Physics 2, more males agreed

that they were recognized by other students as a person that can do physics compared to at the beginning of Physics 2. For the most part, students' perceptions that they were recognized by their peers as a physics person were the same at the beginning and end of the semester.

2. A teacher encouraged me to pursue physics.

Two survey items were related to recognition by teachers as a physics person. The first statement was *A teacher encouraged me to pursue physics*. The distributions of student responses to this statement are presented in Figure 9. In Physics 1, there are no significant gender differences in the responses to this statement on either the pre- or post-course survey [both $p > 0.1$]. Between 50% and 60% of both males and females agree that a teacher has encouraged them to pursue physics. In Physics 2, however, female students are significantly less likely than male students to agree that a teacher has encouraged them to pursue physics, both at the beginning and end of the semester [pre: $Z = -2.83$, $p < 0.01$; post: $Z = -2.99$, $p < 0.01$]. Only about 35% of females agree with the statement, compared to about 45% of males. In neither Physics 1 nor in Physics 2 did the distributions of student responses change between the pre- and post-course surveys [all p 's > 0.3].

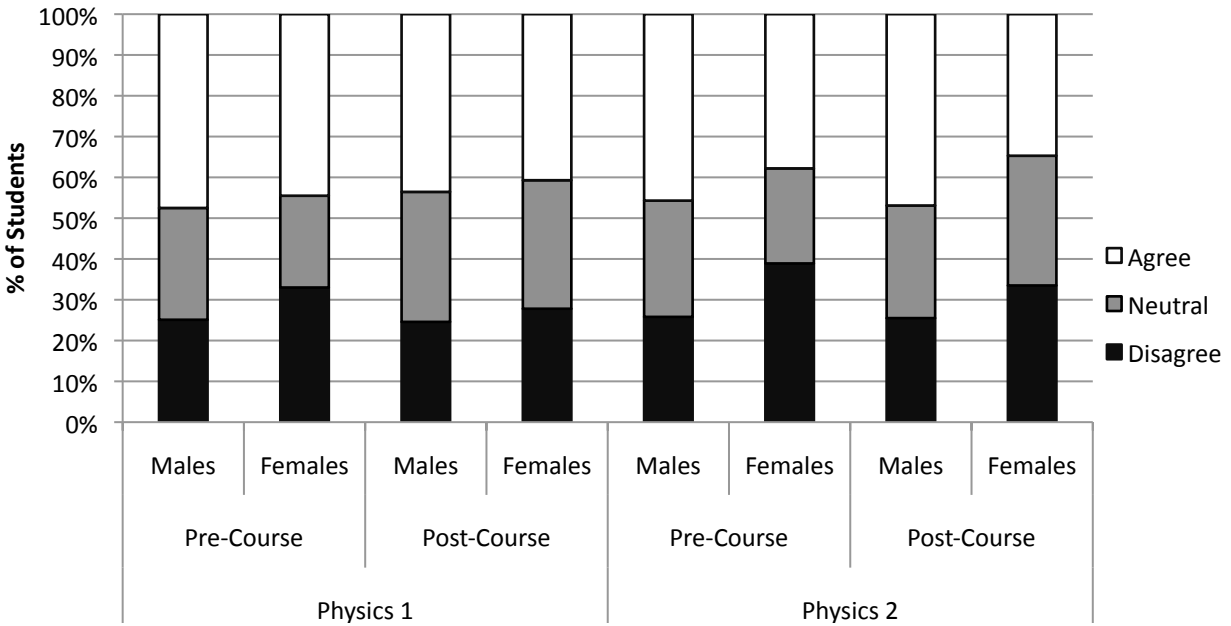


Figure 9. Distribution of male and female responses on the pre- and post-course surveys in Physics 1 and Physics 2 to the statement: *A teacher encouraged me to pursue physics*. The five answer options were collapsed to three categories. In Physics 1, on both the pre- and post-course surveys, there is no significant gender difference in student responses to this statement ($p > 0.1$). There are gender differences on both the pre- and post-course survey in Physics 2 ($p < 0.01$). The distributions of students' responses to this item do not change significantly from pre- to post-course (all $p > 0.3$).

3. A teacher has told me that I'm good at physics.

The second item related to recognition by teachers asked students to agree or disagree with the statement *A teacher has told me that I'm good at physics*. Student responses to this item are presented in Figure 10. At both the beginning and end of Physics 1 and Physics 2, females are significantly less likely than males to agree with this statement [all p 's < 0.01]. Between 35% and 50% of females agree with the statement, compared to between 50% and 60% of males. Unlike the prior item, in which there were not significant differences between students' pre- and post-course responses, for this item, all students shift towards disagreeing with the statement [all p 's < 0.02 , except females in Physics 2 where $p = 0.08$]. Compared to the beginning of the course, students are less likely to agree that a teacher has told them that they are good at physics at the end of the course.

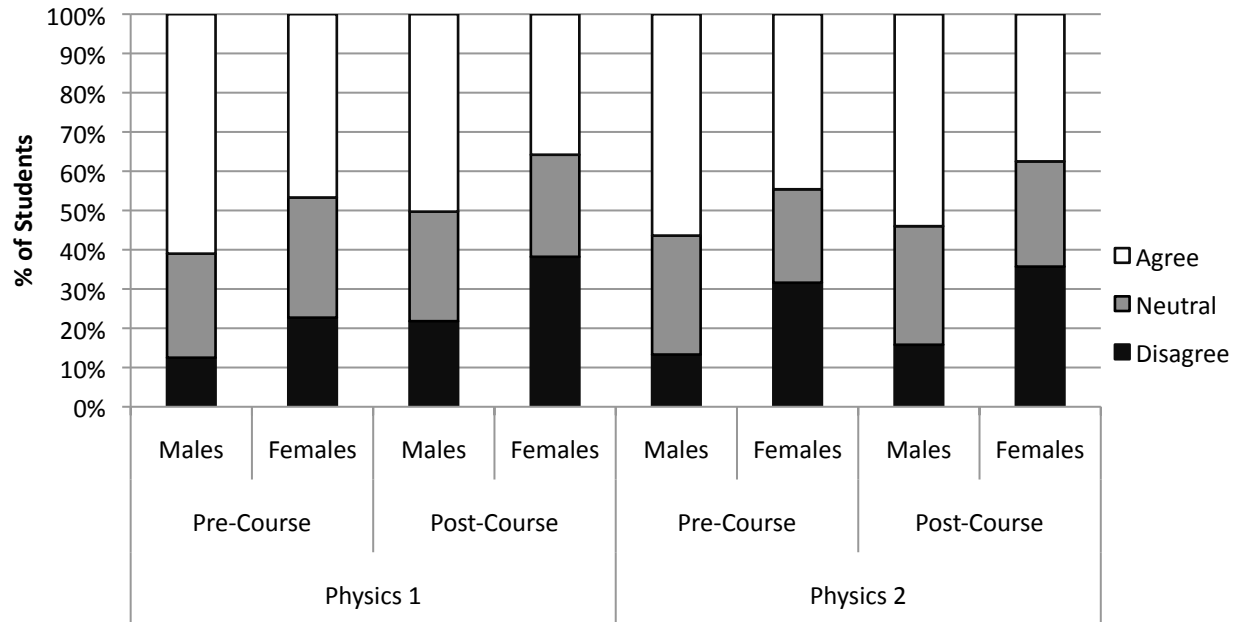


Figure 10. Distribution of male and female responses on the pre- and post-course surveys in Physics 1 and Physics 2 to the statement: *A teacher has told me that I'm good at physics*. The five answer options were collapsed to three categories. In both Physics 1 and Physics 2, on both the pre- and post-course surveys, there are significant gender differences in student responses to this statement ($p < 0.01$). The distributions of students' responses to this item do change significantly from pre- to post-course for all students except females in Physics 2.

4. My parents have encouraged me to pursue physics.

The final item related to students' perceptions that others recognize them as a physics person asks students to agree or disagree with the statement *My parents have encouraged me to pursue physics*. In Physics 1, male and female responses to this statement are not significantly different [pre: $Z = -0.09$, $p = 0.93$; post: $Z = -0.24$, $p = 0.81$]. Between 35% and 40% of males and females agree that their parents encouraged them to pursue physics. Additionally, the distributions of responses do not change from the pre-course to the post-course surveys [males: $Z = -1.77$, $p = 0.08$; females: $Z = -0.74$, $p = 0.46$]. At both the beginning and end of Physics 1, males and females report receiving about equal levels of encouragement from parents to pursue physics.

In Physics 2, however, at both the beginning and end of the semester, females are significantly less likely to agree that their parents encouraged them to pursue physics [pre: $Z=-2.24$, $p=0.03$; post: $Z=-3.19$, $p<0.01$]. Only about 30% of females agree with the statement, compared to about 40% of males. Neither males nor females in Physics 2, significantly change their responses to this item over the semester [males: $Z=-1.39$, $p=0.16$; females: $Z=-0.09$, $p=0.93$].

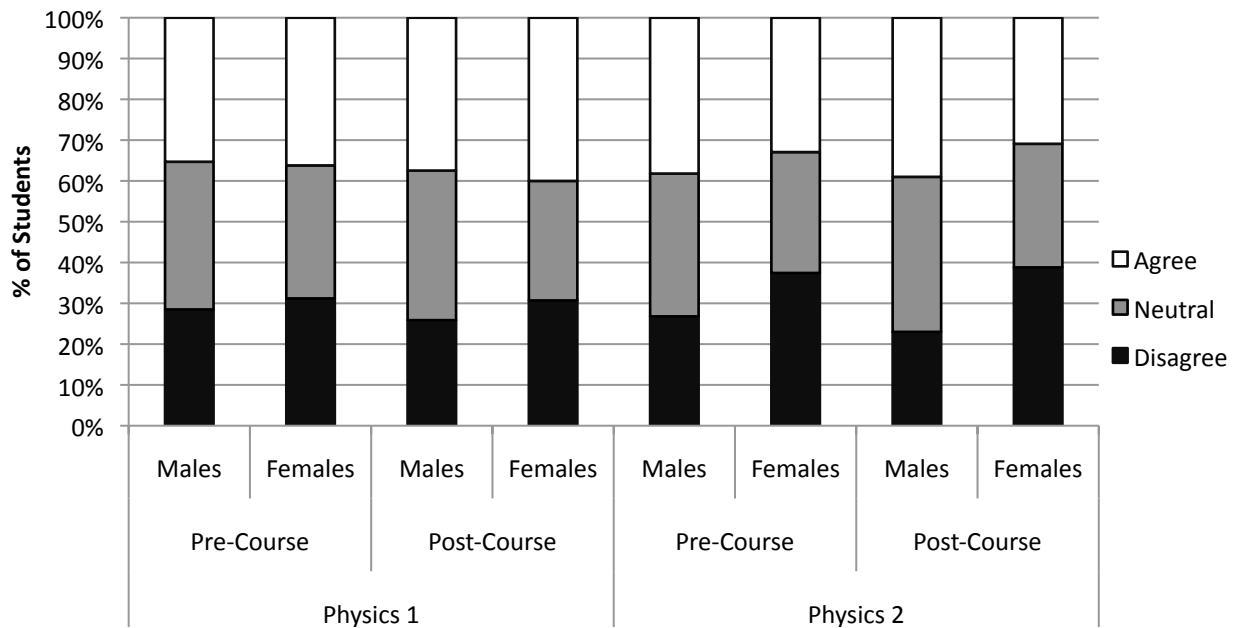


Figure 11. Distribution of male and female responses on the pre- and post-course surveys in Physics 1 and Physics 2 to the statement: *My parents have encouraged me to pursue physics*. The five answer options were collapsed to three categories. In Physics 1, on both the pre- and post-course surveys, there are no significant gender differences in student responses to this statement ($p>0.8$). In Physics 2, female students are significantly more likely to disagree with the statement compared to males on both the pre- and post-course surveys ($p<0.03$). The distributions of students' responses to this item do not change significantly from pre- to post-course for all students.

5. Average Recognition By Others

Just as with self-identity, we look at male and female average values of recognition by others over all statements for each semester of Physics 1 and Physics 2 [26]. These data are presented in Figure 12. In all semesters except the J1-2 semester of Physics 1, females are less

likely to report that other people recognize them as someone that can do physics [all p 's<0.01]. Looking at the shifts in students' reported recognition by others from pre- to post-course, we find that the shifts are not significant in all semesters except the K1-1 semester of Physics 1. For the most part, students' perceptions of peers, teachers, and parents recognizing them as the kind of person that can do physics doesn't change over the course of Physics 1 or Physics 2.

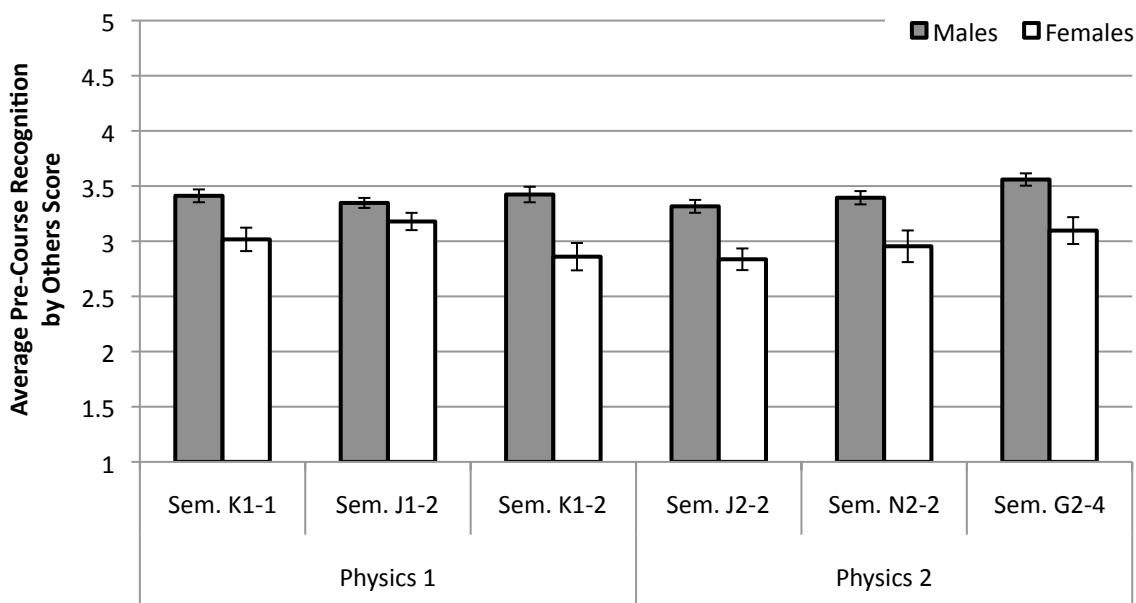


Figure 12. Average pre-course recognition by others scores for males and females in Physics 1 and Physics 2 by semester. In all semesters except the Spring 2010 semester of Physics 1, females report that others are significantly less likely to recognize them as a physics person (all p <0.01).

In the same way that students' self-identity was broken out by students' declared major, we also break recognition by others out by major. The trends for recognition by others are similar to those found for self-identity. There are significant differences in recognition by others as a physics person by students' major [via ANOVA, both p <0.01]. Both male and female physics majors report the highest level of recognition by others. Gender differences exist within almost all majors in both Physics 1 and Physics 2, but the differences are significant only for engineering and other science majors (likely because of the small numbers of students in the

other major categories). Again, accounting for student major explains some but not all of the gender differences in recognition by others that are observed.

C. Regression Modeling

In addition to exploring students' reported self-identity and recognition by others, and gender differences in each, we also want to know whether identification by self or by others is a useful predictor of student performance in Physics 1 and Physics 2. We use a multiple regression analysis to model students' FMCE post-test scores as a function of their gender, prior physics and mathematics understanding, and identity. In addition to the regression model generated to predict FMCE post-test scores, models to predict students' exam scores in Physics 1 and Physics 2 were also developed. The trends for each of these dependent variables were the same, so we present the model of FMCE post-test scores as a representative example. Details of the regression analysis for exams can be found in Appendix B.

Four regression models are presented in Table 1. The independent variables in each model significantly predict students' FMCE post-test scores [all F^2 's > 23.00, p 's < 0.01]. In Model 1, *Female* is the only variable included in the model. The coefficient of this variable indicates that the gender difference in FMCE post-test scores for this sample is 11.28%. In Model 2, covariates are included to control for students' prior physics and mathematics understanding as well as the semester that they took Physics 1. Controlling for these background variables reduces the gender gap from 11.28% to 2.88%. In Model 3, the two variables related to identity are included in the regression model. This model accounts for 41% of the variance in FMCE post-test scores. As can be seen, students' pre-course self-identity is a significant predictor of students' post-course FMCE scores [$\beta=0.09$, $t(725)=2.23$, $p=0.03$]. Whether or not students see themselves as a physics person significantly impacts their performance on an end-of-semester

conceptual survey. Students' pre-course perceptions about whether other people see them as a physics person, however, are not a useful predictor of performance [$\beta=-0.04$, $t(725)=-0.86$, $p=0.39$].

Though self-identity appears to be a useful predictor of student performance according to Model 3, we find that when other independent variables are included in the model, students' self-identity does not remain significant. In Model 4, we include students' pre-course *performance* self-efficacy, which is a significant predictor of FMCE post-test scores [$\beta=0.12$, $t(725)=3.39$, $p<0.01$]. When pre-course self-efficacy is included, students' pre-course self-identity is no longer a significant predictor of end-of-semester performance [$\beta=0.01$, $t(725)=0.24$, $p=0.81$]. While self-identity can be a useful predictor of FMCE performance, other factors, like self-efficacy are more powerful.

Table 1. Coefficient estimates and multiple regression model statistics for each regression model. The dependent variable is the FMCE post-test score. B is the un-standardized coefficient, β is the standardized coefficient.

	Model 1	Model 2	Model 3	Model 4
Model-Level Statistics				
Multiple R^2	0.031	0.403	0.407	0.416
F statistic p value	<0.001	<0.001	<0.001	<0.001
Std. Error	128.4	22.4	22.3	22.2
Independent Variables				
Constant	B 68.2 β <0.01 Sig. <0.01	B 44.7 β <0.01 Sig. <0.01	B 39.03 β <0.01 Sig. <0.01	B 26.05 β <0.01 Sig. <0.01
<i>Female</i>	-11.28 β -0.18 Sig. <0.01	-2.88 β -0.05 Sig. 0.13	-2.19 β -0.03 Sig. 0.25	-1.45 β -0.02 Sig. 0.45
FMCE Pretest	0.53 β 0.48 Sig. <0.01	6.65 β 0.22 Sig. <0.01	0.51 β 0.47 Sig. <0.01	0.49 β 0.45 Sig. <0.01
Math Score	6.62 β 0.11 Sig. <0.01	-1.13 β -0.02 Sig. 0.60	6.96 β 0.12 Sig. <0.01	5.95 β 0.20 Sig. <0.01
Semester J1-2			-0.36 β -0.01 Sig. 0.87	6.79 β 0.12 Sig. <0.01
Semester K1-2				0.36 β 0.01 Sig. 0.87
Pre-Self-Identity			2.59 β 0.09 Sig. 0.03	0.25 β 0.01 Sig. 0.81
Pre-Recognition by Others			-1.08 β -0.04 Sig. 0.39	
Pre-Performance SE			4.71	0.12 Sig. <0.01

V. Discussion and Conclusions

In this study, we explored students' own sense of their physics identity and their perception that they are recognized by others as a physics person. Females report both having a lower sense of physics self-identity and being recognized less by peers, teachers, and parents as a physics person. These differences remain even when controlling for students' declared major, except among physics majors. From students' open-responses to questions about their physics self-identity, students consider interest and competence/performance to be important indicators of whether they consider themselves a physics person. This aligns with our model of physics identity, in that students identified three of the four components as being important to their sense of whether they are a physics person. From the open-responses, we did not find any evidence that recognition by meaningful others was critical for students' assessment of self-identity, as Carlone and Johnson found [20]. This could be due to differences in the populations that were studied. Recall that Carlone and Johnson studied the identities of successful female scientists, whereas our work focuses on introductory college physics students. Our results together with Carlone and Johnson's work, suggest that interest and competence/performance may be important in students' early physics identity development, but to maintain and further develop a physics identity, recognition by meaningful others is critical.

We did not find that students' self-identity or their recognition by others was a significant predictor of student performance in the introductory physics courses when other measures of student background and physics self-efficacy were included. It is still possible, however, that students' sense of physics identity affects their prior physics and mathematics understanding (in terms of the high school courses they took) and their physics self-efficacy (since students' beliefs about their competence in physics is one component of identity), and in that way is important for

students' performance. Additionally, though identity was not a useful predictor of performance, it may prove important, as others have suggested [20,21,22,23], in understanding student retention in physics and why some students leave the physics major.

The negative shifts that we observe over the course of Physics 1 and Physics 2 in students' beliefs that they are the kind of person that can do physics, especially among female physics majors, is troubling. It is important for all students, but especially those students who want to pursue physics as a college major, to see themselves as the kind of person that can do physics, to believe that they are capable of understanding physics and engaging in conversations and arguments where physics understanding is required. Examining students' explanations for why they saw themselves as less of a physics person offers some insight into how the introductory physics course can be changed to promote development of successful physics identities. Most students lost interest in physics or found physics to be too difficult. Incorporating more topics of broad interest in the introductory courses could help to show students how physics connects to their specific, non-physics interests or their lives outside of class. There is some evidence that adopting a curriculum that incorporates topics that girls report being interested in, can, when combined with other techniques that promote girls' engagement in the physics class, prevent a loss of interest in physics among both girls and boys compared to traditional classes [27]. Carlone [15] also identified that the portrayed and perceived difficulty of a reformed high school physics class was a factor that led female students to feel alienated and disconnected with physics. It was made clear to students that not everyone would be able to understand everything that was taught, and there was a perception among students that the people with raw talent in physics were the boys. Regardless of whether these ideas were true (there were girls and boys who both succeeded and struggled in the class), they sent the message

to girls that they were not the kind of person who would be good at physics. Interestingly, she found that other ways in which the difficulty of physics was conveyed were not necessarily based on students' ability, but rather on arbitrary factors, like having to do many problems in a limited amount of time. By changing the messages that are sent to students in the introductory physics courses about who can do physics and removing some of the arbitrary ways in which physics is made difficult, more students may see themselves as people that can do physics.

There are other changes that can be made to the introductory physics courses that can promote the positive development of students' identities. Hazari, *et al.* [21] studied the high school physics experiences that predicted students' physics identity. They found that practices like focusing the class on conceptual understanding, discussing current relevant science, discussing the benefits of being a scientist, and encouraging students to take science classes all positively related to students' physics identity. One other factor that they found to significantly increase females' physics identity (while not hurting males') was to discuss female-underrepresentation in physics. Incorporating one or more these practices in the introductory physics courses could go a long way towards increasing students' physics identity.

If we want to increase the participation of females in physics, we need to focus not only on how students are performing in the course, but also on their development as physicists. This study has shown that females' self-identities, more than males', are negatively impacted by the introductory physics courses, in part because they think they are not doing well in the course and believe they are not capable of doing physics. In the next chapter, we bring together these two ideas, performance and identity. We implement an intervention aimed at alleviating the threat to female students' identities as a way to address gender disparities in performance.

Peer-reviewed publications based on this work

- L. E. Kost, S. J. Pollock and N. D. Finkelstein, “Unpacking Gender Differences in Students’ Perceived Experiences in Introductory Physics” in *Proceedings of the 2009 Physics Education Research Conference*, edited by M. Sabella, C. Henderson and C. Singh (AIP Conference Proceedings, New York, 2009) p. 177.

References (Chapter 5)

- [1] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation* (Cambridge University Press, Cambridge, 1991).
- [2] E. Wenger, *Communities of Practice: Learning, Meaning, and Identity* (Cambridge University Press, Cambridge, 1998).
- [3] M. A. Eisenhart and E. Finkel, *Women's Science: Learning and Succeeding from the Margins* (The University of Chicago Press, Chicago, 1998).
- [4] A. C. Barton, Teaching science with homeless children: Pedagogy, representation, and identity, *Journal of Research in Science Teaching* **35**(4), 379 (1998).
- [5] P. Cobb, Mathematics, literacies, and identity, *Reading Research Quarterly* **39**, 332 (2004).
- [6] A. T. Danielsson and C. Linder, Learning in physics by doing laboratory work: Towards a new conceptual framework, *Gender and Education* **21**(2), 129 (2009).
- [7] J. D. Lee, Which kids can “become” scientists? Effects of gender, self-concepts, and perceptions of scientists, *Social Psychology Quarterly* **61**(3), 199 (1998).
- [8] J. D. Lee, More than ability: Gender and personal relationships influence science and technology involvement, *Sociology of Education* **75**, 349 (2002).
- [9] F. K. Stage and S. A. Maple, Incompatible Goals: Narratives of graduate women in the mathematics pipeline, *American Educational Research Journal* **33**(1), 23 (1996).
- [10] J. Boaler, The development of disciplinary relationships: knowledge, practice, and identity in mathematics classrooms, *For the Learning of Mathematics* **22**(1), 42 (2002).
- [11] K. Tonso, Student engineers and engineer identity: Campus engineer identities as figured world, *Cultural Studies of Science Education* **1**, 273 (2006).
- [12] J. Boaler and J. Greeno, “Identity, agency, and knowing in mathematics worlds” in *Multiple Perspectives on Mathematics Teaching and Learning* edited by J. Boaler (Ablex Publishing, Westport, CT, 2000) p. 171.
- [13] N. W. Brickhouse, P. Lowery and K. Schultz, What kind of a girl does science? The construction of school science identities, *Journal of Research in Science Teaching* **37**(5), 441 (2000).
- [14] N. W. Brickhouse, Embodying science: A feminist perspective on learning, *Journal of Research in Science Teaching* **38**(3), 282 (2001).
- [15] H. B. Carlone, The cultural reproduction of science in reform-based physics: Girls' access, participation, and resistance, *Journal of Research in Science Teaching* **41**(4), 392 (2004).

- [16] G. Hughes, Exploring the availability of student scientist identities within curriculum discourse: An anti-essentialist approach to gender-inclusive science, *Gender and Education* **13**(3), 275 (2001).
- [17] N. W. Brickhouse and J. T. Potter, Young women's scientific identity formation in an urban context, *Journal of Research in Science Teaching* **38**(8), 965 (2001).
- [18] A. C. Johnson, Unintended consequences: How science professors discourage women of color, *Science Education* **91**(5), 805 (2007).
- [19] K. Piatek-Jimenez, Images of mathematicians: A new perspective on the shortage of women in mathematical careers, *Mathematics Education* **40**, 633 (2008).
- [20] H. B. Carlone and A. Johnson, Understanding the science experiences of successful women of color: Science identity as an analytic lens, *Journal of Research in Science Teaching* **44**(8), 1187 (2007).
- [21] Z. Hazari, G. Sonnert, P. M. Sadler and M. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, *Journal of Research in Science Teaching* **47**(8), 978 (2010).
- [22] A. C. Barton and K. Yang, The culture of power and science education: Learning from Miguel, *Journal of Research in Science Teaching* **37**(8), 871 (2000).
- [23] P. W. U. Chinn, Asian and Pacific Islander women scientists and engineers: A narrative explanation of model minority, gender, and racial stereotypes, *Journal of Research in Science Teaching* **39**(4), 302 (2002).
- [24] J. P. Gee, Identity as an analytic lens for research in education, *Review of Research in Education* **25**, 99 (2000).
- [25] For the Wilcoxon Signed Ranks test used throughout this chapter, the Z score that is reported is similar to a typical z -score which reports how many standard deviations from the mean a value is. For our purposes, the Z score indicates how many standard deviations (which are calculated based on the number of students) an observed shift is from zero shift. See W. J. Conover, *Practical Nonparametric Statistics* (John Wiley & Sons, New York, 1999).
- [26] Summary tables of average values of self-identity and recognition by others can be found in Appendix A.
- [27] P. Häussler and L. Hoffman, An intervention study to enhance girls' interest, self-concept, and achievement in physics classes, *Journal of Research in Science Teaching* **39**(9), 870 (2002).

CHAPTER 6: IDENTITY THREAT AND A SELF-AFFIRMATION INTERVENTION IN PHYSICS 1

I. Introduction

In the prior chapters we have found 1) gender differences in students' performance on in-class exams and a pre- and post-test measure of student conceptual understanding, 2) that this gender gap can be largely (and in some cases entirely) accounted for by differences in the preparation of males and females coming into the physics courses, 3) females report feeling less confident than males in their ability to succeed in the introductory physics courses, and 4) females are less likely than males to report that they see themselves as a physics person. Based on these findings, we hypothesized that identity threat may be inhibiting females' performance in the introductory courses. Researchers have demonstrated that identity threat can be alleviated through self-affirmation. In this chapter, we report on a large-scale classroom study that tested the effectiveness of a psychological intervention, called values affirmation, in reducing the gender gap in Physics 1. We answer the following research question: Can the performance differences between males and females be reduced, or even eliminated, with a self-affirmation intervention?

We begin by describing identity threat and its effect on student performance. We then discuss prior literature on self-affirmation and its demonstrated alleviation of achievement differences between groups of students. We summarize the research methods that were used in this study, including the study design, sample of students, the affirmation exercises, the outcome measures that were analyzed, and the statistical methods that were employed. The effects of the

self-affirmation intervention on each of the outcome measures are presented. We find that the self-affirmation intervention was successful in reducing the gender gap both in students' in-class exam scores and their conceptual post-test scores. The reduction in the gender gap was primarily due to affirmed females' increased performance compared to females in a control group. Further, the effect of the intervention was moderated by students' endorsement of the stereotype that men are expected to do better in physics than women. There was a negative relationship between stereotype endorsement and performance on exams and the conceptual survey for females, unless they had completed the values affirmation exercise, in which case they were buffered from the negative impact of stereotype endorsement. Lastly, we report the results from a replication study. We did not find a statistically significant effect of the intervention on the gender gap in the replication study. We discuss the implications of these results on the potential utility of this intervention for addressing gender disparities in introductory physics.

II. Background

Because gender achievement gaps have important educational and societal implications, several attempts have been made to reduce them. For example, in physics, interactive techniques such as *Peer Instruction* [1] and curricular materials, such as *Tutorials in Introductory Physics* [2] and *Context-Rich Problems* [3,4], can reduce the gender gap in college physics classrooms [5]. Larger-scale attempts to reduce the gender gap in physics include restructuring the entire physics course [6,7,8] or introducing mentoring programs focused on women [9]. While some of these attempts have reduced gender gaps [5], interventions in science education have focused mostly on instructional methods. Many have not taken into account social-psychological and cognitive processes that lead to gender differences in performance and learning. One such

process is identity threat. We describe identity threat below, as well as self-affirmation, a proposed method for alleviating identity threat.

A. *Identity Threat*

Identity threat [10,11,12] occurs when a person's self-integrity (his/her sense of being a good and appropriate person) is threatened due to a negative characterization of an aspect of his/her social identity in a particular setting. Individuals have multiple social identities, for example, their sex, ethnicity, race, age, professional identity, social class, etc. When a person realizes that he/she can be negatively judged based on his/her social identity in a particular setting, that individual's identity has been threatened. For instance, a female student experiences identity threat when she becomes aware that she can be negatively judged in her physics course because she is a woman. An example of identity threat is stereotype threat [10,13,14], where the negative characterization is a specific stereotype about one's group, for instance "women can't do math." But a negative stereotype isn't the only way to trigger identity threat. Any cue that alerts an individual that he/she could be devalued based on his/her group identity can arouse identity threat. Examples of such cues include a lack of people in a setting who share a particular social identity (fewer women in a physics course), or cues that social identity plays an organizational role in the setting (men do better in the physics course) [10]. When people pick up on these cues, they become more aware that they can be negatively judged based on their social identity, and are thus subject to identity threat.

Identity threat has been shown to undermine the performance of stereotyped or threatened students [10,15,16,17]. Spencer, *et al.* [15] demonstrated the effect of stereotype threat on females' mathematics performance. They found that when students were told that a difficult standardized math test generally produced gender differences (thereby arousing stereotype threat

among female students), females performed worse than equally qualified males. However, when students were told that the same test generally did not show gender differences (alleviating females' concerns that their poor performance would confirm the stereotype that "girls can't do math"), females' performance increased, matching that of the males. The negative impact of stereotype threat on females' mathematics performance has been demonstrated across a variety of contexts [18,19,20,21] and for students as young as 5 and 6 years old [22].

The effects of stereotype threat on performance have also been shown for other minority student groups, such as African American and Latino students compared to White students [13,14,17,23]. But it is not only minority students that can be impacted by identity threat. In one study, researchers found that White males underperformed on a difficult mathematics test compared to Asian males [24]. In this study, researchers aroused stereotype threat simply by telling participants that they were studying why Asian males were so good at mathematics. This study demonstrates that students don't need real-life exposure to the stereotype (since there is not a pervasive stereotype that White males are not good at math) or even to believe the stereotype in order to be subject to its threat. Further, identity threat has been shown to affect performance across a variety of performance domains, not just in academic achievement. Researchers have shown that identity threat can undermine athletic performance of Whites and women [25,26], negotiating performance of women [27], memory performance in elderly people [28], and the driving performance of women [29].

Though one need not believe the stereotype or negative characterization of his/her identity in order to be subject to its negative effects, researchers have found that identity threat can be heightened if an individual thinks that the stereotype or negative characterization might be valid or worries that it could be applied to them [30,31]. Schmader *et al.* [30] found that the

effects of stereotype threat were moderated by the degree to which females endorsed the stereotype that women are worse than men at math. That is, females who endorsed the stereotype underperformed on a mathematics test compared to females who did not endorse the stereotype. Other factors that have been shown to moderate stereotype threat are people's theories of intelligence [32,33], whether they believe that intelligence is fixed or malleable, how much a person identifies with their social identity group, gender identification for instance [34], and how much a person identifies with the domain they are being tested in [25,35].

B. *Self-Affirmation*

Several researchers have found that stereotype threat can be alleviated through self-affirmation [36,37]. Self-affirmation is a process whereby a person affirms their overall sense of self-worth and integrity [38,39]. Work on self-affirmation assumes that people are motivated to maintain a positive sense of overall integrity, identity, and worth. When our integrity or identity is threatened, we seek ways to resolve the threat. Because it is often difficult (or even impossible) to resolve the specific identity threat (that "girls can't do science"), individuals can affirm a more general sense of worth and integrity ("I'm a good person") or a specific, but unrelated, aspect of their identity ("I'm good at music"), which will help to protect them from the threat. A common self-affirmation process used in the psychological literature is values affirmation, in which people reflect on self-defining values. This activity focuses individuals' attention on values important to them that are not connected to the threat, allowing them to reestablish a perception of personal integrity and worth and to realize that their self-integrity does not depend on their performance in the threatened situation.

In a laboratory experiment with college students, Martens, *et al.* [36] found that females who were given the opportunity to write about a characteristic of themselves that they valued

performed better on a subsequent difficult math test than females who wrote about a characteristic that they did not value. Similar results were found by Cohen, *et al.* [37] comparing the school achievement of middle school African Americans who did and did not affirm their personal values. Though identity threat has been shown to affect performance across a wide variety of situations and performance settings (including in both laboratory experiments and classroom studies), attempts to reduce identity threat in authentic classroom contexts have been limited [33,37]. Further, the impact of self-affirmation on the gender gap in college physics (using objective measures of student performance) has not been examined.

III. Research Methods

To test the effectiveness of the values affirmation in reducing the gender gap in Physics 1, we conducted a randomized, controlled study. In this section, we describe the experimental design of the study, the course that we conducted the study in, the sample of students analyzed, the experimental (values affirmation) and control writing exercises, the measure of stereotype endorsement, the outcome measures analyzed, and the statistical methodology.

A. Study Design

To test whether a self-affirmation intervention can reduce the gender gap in physics, we used a randomized, controlled study design. Students were randomly assigned to complete one of two brief writing exercises, twice during the semester. Each student completed either a values-affirmation writing intervention or a control writing exercise of similar format and length (the content of the writing exercises will be described below). The writing assignments were delivered early in class (week 1) to maximize the effect of the intervention across the semester, with a second administration shortly before the first exam (week 4) to ensure the potency of the intervention.

The first writing exercise was presented during the first recitation section. During each of the two lectures that students attended prior to the first recitation section, the professor provided pedagogical context for the writing exercise that they would complete. Specifically, he told the students that effective communication was an important skill for success in physics-related careers and that, to practice communication, they would complete a 10 to 15 minute writing exercise in recitation. The professor told them that they would not be writing about physics, but about something that they already knew about.

Recitation sections were led by graduate TAs who were naïve to the purpose of the study. Student attendance was mandatory in these sections, and all recitation sections met on the same day. TAs were given a scripted introduction explaining to the students that they would be completing a writing exercise in which they would think about values that are important to people (TA scripts can be found in Appendix A). To further standardize administration, TAs were given scripted answers to possible questions from the students. They distributed to each student a manila envelope containing the informed consent form and writing assignment that was prepared in advance by study personnel. Although there were two versions for the writing assignment (values affirmation and control), the envelopes and formatting of the two exercises looked similar. Additionally, the TA established the expectation of silent concentration. These steps served to minimize the possibility of students' becoming aware of differences in the exercises. The consent form described the study as examining the relation between critical writing and students' experiences in college, including physics classes. All students over the age of 18 were invited to participate. Students who were younger than 18 and hence could not legally grant consent were instructed to complete an alternative writing assignment, of comparable length, asking about past physics experiences.

In each recitation section, 60% of the packets given to each TA contained the values-affirmation writing assignment. We overrepresented the affirmation condition so that any possible benefits could be conferred to the greatest number of students without undermining the rigor of the study. The affirmation and control packets were intermixed randomly. Random assignment to condition occurred when students received a packet. The two writing exercises were formatted identically and were of nearly identical length. Students put all materials (the writing exercise and the consent form) back in the manila envelope when they were done. TAs collected the envelopes after 15 minutes and were instructed to not open them at any time.

Study personnel discreetly monitored the administration of the writing exercise in each recitation from the back of the classroom. They verified that all TAs properly instructed and administered the exercise. Study personnel discreetly collected the completed materials at the end of each recitation after students had departed.

A second administration of the writing exercise was delivered shortly before the first midterm exam (week 4) via a regular, weekly online homework assignment. Students received the assignment on a Friday, 11 days before the first midterm exam, and could complete it any time within the next 8 days, with a Saturday 8 AM deadline. Consequently, students could have completed the second writing exercise between 3 and 11 days before the first midterm. The majority of students (76%) completed the exercise toward the end of the homework period (the following Thursday, Friday, or Saturday morning). The homework assignment was delivered and completed online. The final question on the homework assignment asked students to follow a link to the university-wide online course portal. Doing so allowed us to customize the content delivered, ensuring that each student received the writing exercise for the same condition that he or she had completed in recitation in week 1. The writing exercise was similar to the one

completed in recitation, but the second exercise was presented online, with students typing their answers and submitting them via the online portal. Instructions suggested that students spend about 15 minutes on the exercise.

Several steps were taken to ensure that all instructional personnel associated with the course were unaware of students' condition assignment. All but the course instructor were blind to the study's purpose and hypotheses. Both writing exercises occurred without the course instructor present (in TA-led recitations and online). All data were handled only by study personnel. The course instructor had no access to information about students' assignment to experimental condition (values affirmation or control) and was not told of any results until after the semester was over. The TAs who distributed the writing exercises in recitation were told only that they would be administering a writing exercise and were not told of the hypotheses or even of the presence of two different writing exercises. The affirmation and control exercises given in recitation were distributed in closed envelopes, with study personnel present in every recitation to verify that TAs never viewed the students' responses. The writing exercises distributed in recitation were formatted identically, preventing TAs from noticing condition assignment. TAs were also instructed to remain at the front of the room while students completed the exercise in recitation. Completed writing exercises were collected by study personnel immediately at the end of each recitation.

Because the writing exercise portion of the homework assignment (week 4) was not required and students were allowed to opt out, there was some attrition across the two administrations of the intervention. Of the total students completing the first writing exercise in week 1, approximately 74% (439 students) completed the second writing exercise in week 4 and took the final exam (thus comprising the students in the study sample). There was a tendency

that, among students who completed the first administration of the intervention, more students completed the second administration in the affirmation condition than in the control condition [$\chi^2(1, N=591)=3.30, p=0.07$], a trend that was almost significant for females [$\chi^2(1, N=158)=3.70, p=0.054$] but not for males [$\chi^2(1, N=433)=1.31, p=0.25$].

B. *Course Description*

We conducted this study in the first-semester, calculus-based introductory mechanics course (Physics 1) at the University of Colorado at Boulder. The course content and pedagogy of this course was in line with prior offerings of the course (as described in detail in Chapter 2). The (male) instructor is a highly experienced teacher, well versed in the interactive engagement techniques that he used during the lecture sections. Typical for this course, approximately $\frac{3}{4}$ of the enrolled students were male (74%). There were 6 Teaching Assistants (5 males and 1 female) for the course, who led the recitation sections, and they were assisted by 7 undergraduate Learning Assistants (4 males and 3 females) [40].

C. *Study Sample*

The study sample consisted of 439 students (311 males and 128 females) who (a) were over 18 years of age, (b) completed both writing exercises in weeks 1 and 4, and (c) took the final exam at the end of the semester. This sample represented 72.9% of the 602 students over 18 years of age who received a grade in the course. Of those 439 students in the sample, 7 students were removed from the analysis for failure to follow the writing instructions for one or both of the writing exercises (e.g., by discussing the personal importance of a value when in the control condition). Twenty-three students did not complete the stereotype endorsement survey question, and 8 students did not have prior mathematics performance data (either SAT- or ACT-Math scores). An additional 2 students were removed based on two multivariate outlier analyses [41],

one involving the four exam scores and one involving the conceptual survey scores obtained at the beginning and end of the semester. These students had Mahalanobis distances [42] of 22.6 for the exam scores (critical value=18.5) and 14.0 for the conceptual survey scores (critical value=13.8), respectively.

This left a final sample of 399 students (283 males and 116 females) on which the reported analyses were based. An analysis of the attrition from the original study sample (439 students) to the final sample (399 students) showed no difference in attrition by condition for all students [$\chi^2(1, N=439)=1.35, p=0.24$], or for females specifically [$\chi^2(1, N=128)=0.01, p=0.94$].

As noted above, we randomly assigned approximately 60% of students to the values affirmation condition and 40% to the control condition. This assignment plan was largely successful; the final sample of 399 students consisted of 178 males and 69 females in the affirmation condition (62.9% of males and 59.5% of females in the final sample) and 105 males and 47 females in the control condition (37.1% and 40.5%, respectively). Because 308 (77.2%) of the 399 students (212 males and 96 females) in the final sample took the conceptual survey in both weeks 1 and 15, the sample used for the analysis of the conceptual scores consisted of 137 males and 55 females in the affirmation condition (64.6% of males and 57.3% of females in this sample) and 75 males and 41 females in the control condition (35.4% and 42.7%, respectively). The numbers of males and females in the affirmation and control conditions for each of the samples used in this study can be found in Table 1.

Table 1. Numbers of males and females in the affirmation and control conditions for each of the samples used in this study.

	Males	Females
Final Study Sample (N=399)		
Affirmation	178	69
Control	105	47
Conceptual Survey Sample (N=308)		
Affirmation	137	55
Control	75	41

To ensure that there were no background differences between the affirmation and control groups, we compared several measures of student background for males in the affirmation versus control conditions, and similarly for females. Twelve measures of student background were compared (e.g., high school GPA, SAT and ACT total scores and each component score, and high school courses taken; a complete list is given in Ref. 43) and none were significantly different by condition (all p 's > 0.09). This ensures that, at least for the background variables that we had access to, students in the affirmation and control conditions were similar coming into the Physics 1 course.

D. *Values Affirmation and Control Writing Exercises*

The values affirmation and control exercises closely followed procedures developed and validated in prior research [37,44]. Both writing exercises can be found in Appendix A. Students in each writing condition received a three-page packet. The first page listed 12 values: *being good at art; creativity; relationships with family and friends; government or politics; independence; learning and gaining knowledge; athletic ability; belonging to a social group (such as your community, racial group, or school club); music; career; spiritual or religious values; and sense of humor*. The values were similar to those used in past research [37,44], though modified somewhat for the present sample [45], and were selected to represent a range of

values that students may or may not endorse. We avoided values that explicitly dealt with science and math, so that students were forced to affirm a value unrelated to the threatened domain. Students in the affirmation condition were instructed to circle the two or three values *most* important to them, whereas students in the control condition were instructed to circle the two or three *least* important values.

Through a series of structured prompts, the second page of the packet instructed students to describe in a few sentences either why the selected values were important to *them* (affirmation condition) or why they might be important to *someone else* (control condition). To decrease evaluation apprehension, students were told to focus on their thoughts and feelings, without worrying about spelling and grammar or how well written their answer was. Lines were provided on two thirds of the page for students to provide their answer.

The final page reinforced the manipulation by asking students to again look at the values they had selected earlier. They were then asked to list either the top two reasons why these values were important to them (affirmation condition) or the top two reasons why these values might be picked as important by someone else, such as another student at their school or a person they have heard about (control condition). To further encourage reflection about the values, the third page ended by asking students to indicate their agreement with several items using numerical scales (e.g. *In general, I try to live up to these values* in the affirmation condition vs. *In general, some people try to live up to these values* in the control condition).

E. *Stereotype Endorsement Measure*

In the second week of the course, students were asked to complete a survey about their attitudes toward science (the survey can be found in Appendix B). This survey is a typical part of this introductory physics course. Embedded in the larger survey was an item asking them to

report their expectation that men do better in physics than women. The survey was included as a link on their weekly online homework assignment. Students were asked to follow the link to complete the survey and were told that they would receive extra credit (equivalent to one homework problem) if they included their name on the survey (with no requirement that they answer any questions). Clicking on the link brought students to an online survey containing various attitude measures (e.g., the Colorado Learning Attitudes about Science Survey [46]). Stereotype endorsement was measured by assessing students' agreement with the statement, *According to my own personal beliefs, I expect men to generally do better in physics than women*, answered on a 5-point scale ranging from *strongly disagree* to *strongly agree*. The distributions of male and female responses to this statement are presented in Figure 1. Females were significantly more likely than males to disagree with this statement [$Z=-6.00, p<0.001$].

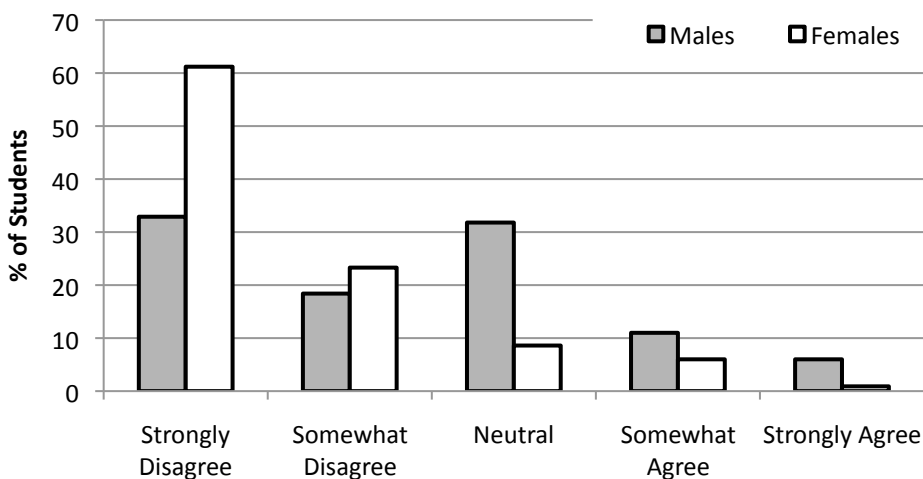


Figure 1. Distributions of male and females responses to the statement: *According to my own personal beliefs, I expect men to generally do better in physics than women*. Females are significantly more likely than males to disagree with this statement ($p<0.01$).

Because the measure of stereotype endorsement was collected in week 2, after the first administration of the writing exercise, we evaluated whether the writing exercise affected

stereotype perceptions. Neither the *condition* main effect nor the *gender* × *condition* interaction was significant [$F(1,395)=0.05$, $p=0.82$ and $F(1,395)=0.31$, $p=0.58$, respectively], suggesting that there was no effect of the affirmation versus control writing exercises in week 1 on students' reported endorsement of the stereotype in week 2.

F. *Outcome Measures*

Several outcome measures were used in this study to test the impact of the values affirmation intervention. The main outcome measures were students' exam scores and the Force and Motion Conceptual Evaluation (FMCE) [47] post-test scores, since these are the measures where we observed the largest gender differences (see Chapters 2 and 3). Students took three midterm exams and a final exam in this semester. All exams were multiple-choice, with objectively correct versus incorrect answers, and were machine-graded rather than graded by the instructor. The final exam was cumulative. The FMCE was administered to students both at the beginning and at the end of the semester. Students took the FMCE pre-test during the first week of the course in recitation. After students completed the first writing exercise for the values affirmation or control condition, they completed the FMCE for the remainder of the recitation section (approximately 35 minutes). The post-course administration of the FMCE took place during the last week of the semester, again in recitation. Students were given the entire 50 minutes to complete the post-test. In both administrations, students were explicitly told that the FMCE scores would not affect their course grades in any way.

Students' homework, participation, and clicker scores were also examined, as well as students' final course grades. Like in other semesters, students' completed two types of homework assignments, CAPA [48] homework, which was administered online and was made up of traditional "end-of-chapter" problems, and *Tutorial* [2] homework, which was associated

with their recitation activities. We examined students' scores on both types of homework separately. Participation in the course was determined by whether students completed the online *Tutorial* pretests and whether they attended and participated in the recitation sections. As in other semesters, the professor interspersed the lecture with conceptual questions. Students responded to the questions using personal response devices (clickers). Students received some credit just for answering the clicker questions and received more credit if they answered the questions correctly. Points that students earned by answering clicker questions were only counted as extra credit towards students' grades, but we examined their clicker scores nonetheless.

Students' overall course scores were composed of their exam, homework, and participation scores. Each of the midterm exams accounted for 14% of the final score (a total of 42%), the final exam for 33%, and the homework and participation for the remaining 25%. Students' final course scores were first computed as a percentage. Based on the distribution of these course score percentages, the professor assigned course grades (A, B, C, etc.).

G. *Statistical Methodology*

The effectiveness of values affirmation in reducing the gender gap was assessed by conducting separate multiple regression analyses [49] for each outcome measure. Of particular theoretical interest in each analysis was the interaction between student gender and condition, which tested whether the performance of females in the affirmation condition was improved relative to females in the control condition and whether the gender gap was reduced in the affirmation condition. We were also interested in the three-way interaction between gender, condition, and stereotype endorsement, which tested whether the *gender* \times *condition* interaction varied depending on students' level of agreement with the stereotype that men do better in physics than women.

To test for these critical effects, we conducted a series of regression analyses and included the following predictors in the models: subject gender (1=females, -1=males), condition (1=values affirmation, -1=control), and stereotype endorsement (mean centered for all students, following reference 30), plus all two-way interactions (*gender* × *condition*, *gender* × *stereotype endorsement*, and *condition* × *stereotype endorsement*) and the three-way *gender* × *condition* × *stereotype endorsement* interaction.

When assessing the effects of identity threat on performance, it is critical to evaluate these theoretically predicted effects while controlling for prior relevant performance [17,50]. Because previous research has shown that background preparation in math predicts physics grades [51] and accounts for a substantial amount of variance in gender differences in performance in physics (see Chapter 2 and 3), a mean-centered measure of prior math background was included as a covariate in the analyses (Math Score). This measure was calculated first by standardizing students' SAT- and ACT- Math scores (provided from university records) and then using whichever of the two scores was available or, in cases where scores for both tests were available from student university records, using the average of the two. The analysis of the FMCE post-test scores included the same predictors, except that the FMCE pre-test score (mean centered) was used as a covariate instead of prior math background.

As is recommended [52] when background variables correlate with variables of interest (as Math Score and FMCE pre-test both correlate with gender), to avoid biases in the estimation of regression coefficients, we include terms representing the interaction of the background variables (Math Score or FMCE pre-test score) with gender, condition, and stereotype endorsement. Specifically, the regression models we present in this chapter contain the following 11 predictors: *gender*; *affirmation condition*; *stereotype endorsement*; *gender* × *condition*;

gender × *stereotype endorsement*; *condition* × *stereotype endorsement*; *gender* × *condition* × *stereotype endorsement*; *math score*; *math score* × *gender*; *math score* × *condition*; and *math score* × *stereotype endorsement*. The analysis of FMCE post-test scores is similar, but uses FMCE pre-test scores as the covariate.

Because these interaction terms involving the covariates were included to ensure that the tests of our predicted effects were unbiased, and because those interaction terms themselves are not of direct theoretical relevance to the understanding of gender identity threat effects, results involving the covariates are not discussed. We also evaluated more complex models that included all possible 3-way and 4-way interactions with the background covariate variables (e.g., the *math score* × *gender* × *condition* interaction in the analyses of exam scores), but in no case were any 3-way or 4-way interactions significant. Thus, for simplicity, we report here the analyses that included only the 11 predictors described above.

Throughout the Results section, we present the standardized (β) coefficients resulting from the regression analyses. All reported p levels are two-tailed.

IV. Results

In this section, we report the effect of the intervention on each of the outcome measures that were discussed above. For the exams and FMCE, we predicted a reduced gender gap in performance for females who completed the values affirmation. Moreover, because people who endorse negative stereotypes about their group are most vulnerable to identity threat [30], we expected the intervention to be particularly beneficial for females tending to endorse the gender stereotype. A table of the raw and adjusted scores for each of the outcome measures can be found in Appendix C.

A. Exams

1. Mean Overall Exam Scores

An overall mean exam score, created by averaging the percent correct on each of the three midterms and the final exam, was used as the primary dependent measure for exam scores. Students' overall mean exam score was regressed on the eleven predictor variables listed above. The results of the regression analysis are presented in Table 2.

Table 2. Multiple regression coefficient estimates and significance levels. The dependent variable is the mean overall exam score. The independent variables significantly predicted students' mean overall exam scores [$F(1,387)=16.00, p<0.01$]. This model accounts for 31% of the variance in mean overall exam score ($R^2=0.31$).

Variables	β	$t(387)$	Significance
Gender	-0.23	-4.62	<0.01
Condition	0.05	0.94	0.35
Stereotype Endorsement	-0.13	-2.31	0.02
Math Score	0.43	9.00	<0.01
Gender \times Condition	0.16	3.08	<0.01
Gender \times Stereotype Endorsement	-0.06	-1.09	0.28
Gender \times Math Score	-0.03	-0.57	0.57
Condition \times Stereotype Endorsement	0.15	2.64	<0.01
Condition \times Math Score	0.08	1.71	0.09
Stereotype Endorsement \times Math Score	-0.10	-2.08	0.04
Gender \times Condition \times Stereotype Endorsement	0.16	2.74	<0.01

Replicating our past research on gender gaps in physics, there was a significant gender difference in students' exam scores in this semester [$\beta=-0.23, t(387)=-4.62, p<0.01$], with males ($M=70.7\%$) scoring higher than females ($M=64.2\%$) across the four exams. This gender difference was present even when controlling for differences in prior math performance. This was the case for all students and for only those students in the control condition. As shown in Figure 2, males outperformed females in the control condition on overall exam scores. However, the gender gap was significantly smaller in the affirmation condition than in the control condition, resulting in a significant *gender \times condition* interaction [$\beta=0.16, t(387)=3.08, p<0.01$].

The effect size for the observed gender gap was substantial in the control condition (Cohen's $d=0.93$ [53]) [$F(1,387)=36.71, p<0.01$], but much smaller, and not statistically significant, in the affirmation condition ($d=0.18$) [$F(1,387)=2.35, p=0.13$]. Further examining the simple effects [49], we find that for females, affirmation resulted in higher exam scores compared to those in the control condition [$F(1,387)=7.61, p<0.01$]. There was also an unexpected yet significant tendency for males to have lower exam scores in the affirmation than control condition [$F(1,387)=5.37, p=0.02$]. This finding will be discussed in more detail below.

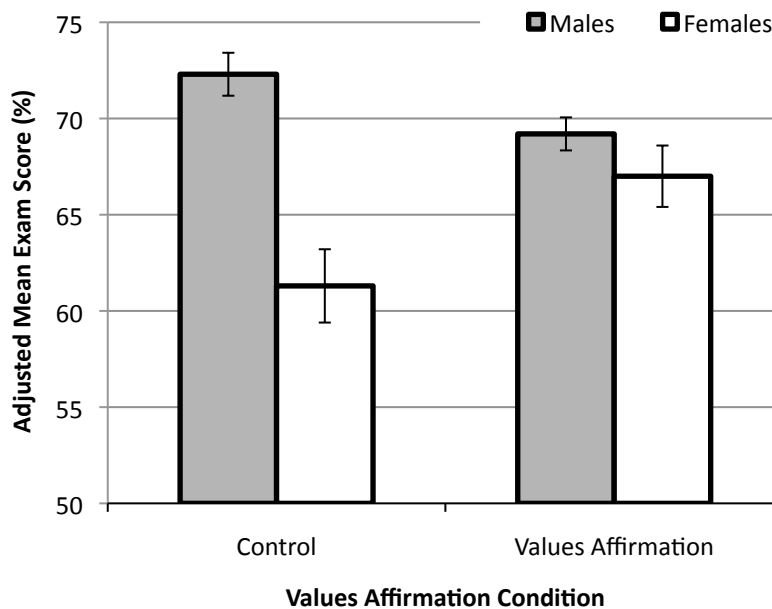


Figure 2. Student overall mean performance on exams as a function of gender and condition. The mean exam score is adjusted based on students' prior math score. The error bars represent the standard error on the mean. The gender difference in the control condition is statistically significant ($p<0.01$), but the gender gap in the affirmation condition is not statistically significant ($p=0.13$).

In addition, we find that the values affirmation was particularly beneficial for females who tended to endorse the gender stereotype. This moderation effect is illustrated in Figure 3. Although females as a group did not strongly endorse the negative gender stereotype (see Figure 1) [30], even a moderate level of stereotype endorsement was costly for females in the control condition, with their exam scores decreasing as a function of stereotype endorsement [$\beta=-0.50$,

$t(387)=-3.29, p<0.01$]. Affirmation, however, buffered females against this identity threat, eliminating the negative relation between stereotype endorsement and exam scores [$\beta=0.12, t(387)=0.94, p=0.35$]. Moreover, among females expressing higher levels of stereotype endorsement (defined as 0.75 SDs above the mean here), affirmation improved the exam scores relative to the control condition [$t(115)=3.04, p<0.01$]. In contrast, males' exam scores were little affected by stereotype endorsement, regardless of condition [$\beta=-0.08, t(387)=-1.70, p=0.09$, for the affirmation condition, and $\beta=-0.07, t(387)=-0.92, p=0.36$, for the control condition]. These differential patterns for males and females resulted in a significant *gender* \times *condition* \times *stereotype endorsement* interaction [$\beta=0.16, t(387)=2.74, p<0.01$].

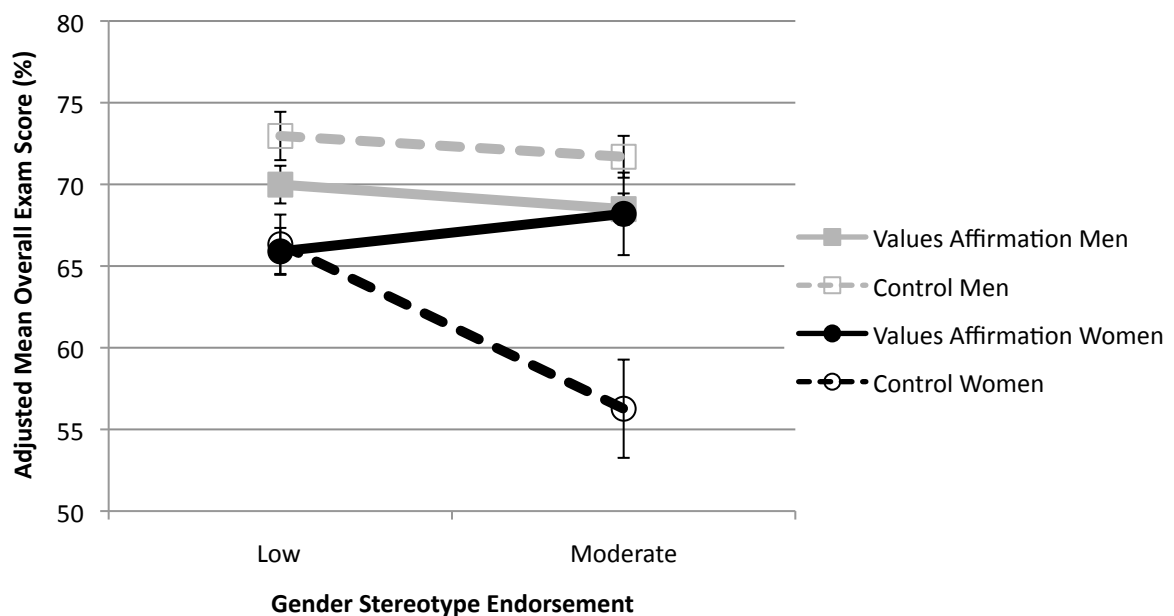


Figure 3. Student overall mean performance on exams as a function of gender, condition and the level of stereotype endorsement. The mean exam score is adjusted based on students' prior math score. The error bars represent the standard error on the mean. Moderate and low gender stereotype endorsement represent ± 0.75 SD of the grand mean of stereotype endorsement. The relationship between mean exam score and stereotype endorsement is only significant for females in the control group. Among females who moderately endorsed the gender stereotype, females in the control group had significantly lower mean exam scores than females in the affirmation group.

2. Individual Exam Scores

In addition to looking at students' overall mean exam scores, we can also examine the scores on individual exams. We find that the affirmation intervention was successful at reducing the gender gap on three of the four exams. As in Figure 4, on the first, second, and final exams, the gender gap in the affirmation condition is significantly smaller than the gender gap in the control condition [all *gender* × *condition* interactions significant, $t's(387) > 2.20$, $p's < 0.03$] (regression results for each exam can be found in Appendix D). Only on exam three were the gender gaps in the affirmation and control groups not statistically different [*gender* × *condition*: $\beta = 0.08$, $t(387) = 1.48$, $p = 0.14$]. For exams one, two, and the final, females in the affirmation condition scored significantly higher than females in the control condition [$F's(1,387) > 4.90$, $p's < 0.03$]. Males in the affirmation and control conditions did not have significantly different scores [$F's(1,387) < 3.65$, $p's > 0.05$]. Additionally, the *gender* × *condition* × *stereotype endorsement* interaction was significant for the first, second, and final exams [$t's(387) > 2.10$, $p's < 0.04$]. The results that we saw for the mean overall exam score also held for three of the four individual exams. Although the second affirmation exercise was completed shortly before the first midterm exam, and no further exercises were completed the rest of the semester, the benefits of the intervention remained evident throughout the semester.

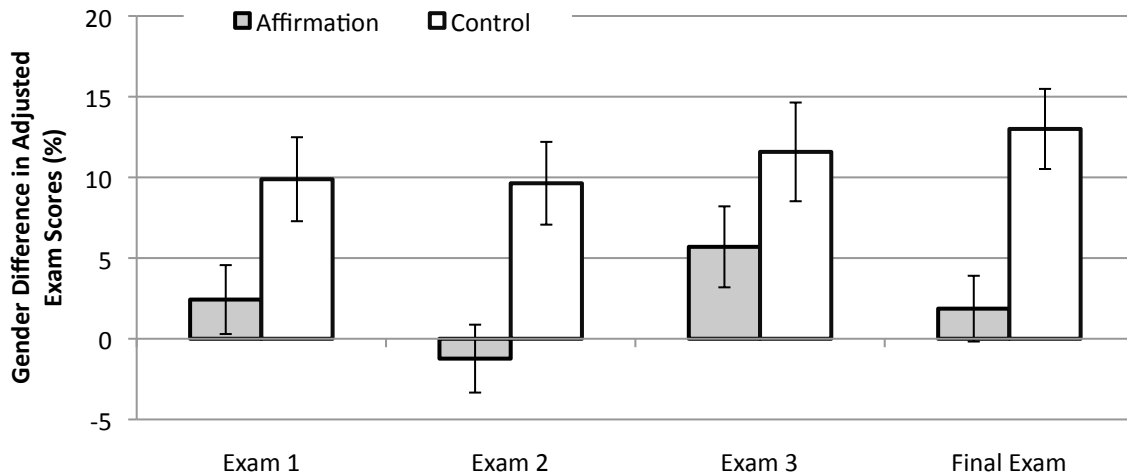


Figure 4. Gender differences on each exam as a function of condition. Exam scores are adjusted based on students' prior math score. The error bars represent the standard error on the mean. For the first, second, and final exams the gender gap among students in the affirmation condition is significantly smaller than the gender gap in the control condition. For exam 3, there is no statistically significant difference in the gender gaps for the affirmation and control conditions.

B. Course Grade

We examine both students' course score (the percentage of total points that they earned) and their letter grade. Students' course scores showed a pattern of effects highly similar to the exams. This is not surprising, given that 75% of the final course grade was based on the four exams. The average final course score was higher for males ($M=74.7\%$) than females ($M=70.0\%$) [$\beta=-0.18$, $t(387)=-3.70$, $p<0.01$], even after controlling for differences in prior math background. The predicted *gender* \times *condition* interaction was also significant [$\beta=0.14$, $t(387)=2.50$, $p=0.01$]. For females, the affirmation improved their course score relative to females in the control condition [$F(1,387)=5.26$, $p=0.02$], whereas the affirmation versus control difference was marginally significant for males [$F(1,387)=3.23$, $p=0.07$]. This resulted in a significant gender gap in the control condition [$F(1,387)=23.82$, $p<0.01$], which was eliminated in the affirmation condition [$F(1,387)=1.43$, $p=0.23$].

As with exam scores, the benefit of values affirmation on course scores was moderated,

as predicted, by stereotype endorsement. The two-way *condition* × *stereotype endorsement* was significant [$\beta=0.15$, $t(387)=2.51$, $p=0.01$] as well as the predicted higher-order *gender* × *condition* × *stereotype endorsement* interaction [$\beta=0.16$, $t(387)=2.69$, $p<0.01$]. Decomposing the three-way interaction, we found that there was no relationship between stereotype endorsement and course score for males in either the affirmation or control condition [$\beta=-0.06$, $t(387)=-1.46$, $p=0.14$, and $\beta=-0.04$, $t(387)=-0.56$, $p=0.58$, respectively]. For females, however, there was a significant negative relationship between stereotype endorsement and course score in the control condition [$\beta=-0.46$, $t(387)=-2.96$, $p<0.01$], but not in the affirmation condition [$\beta=0.15$, $t(387)=1.23$, $p=0.22$]. Among the females relatively high in stereotype endorsement (0.75 *SDs* above the mean), the course score was significantly higher in the affirmation condition than in the control condition [$t(115)=2.74$, $p<0.01$].

Examining the letter grades that students' received, we find that the values affirmation was particularly beneficial in elevating females' course grades from average to above average. As can be seen in Figure 5, a large majority of females in the control condition (55.8%) earned a grade in the C range (including C-, C, and C+), with only 23.1% earning a grade in the B range (including B-, B, and B+). The percentage of C's was reduced to 40.8% among females in the affirmation condition, and B's increased to 36.8%. This difference in the percentage of females getting B's and C's across the two conditions was statistically significant [$\chi^2(1, N=91)=4.07$, $p=0.04$]. There was no difference in the distribution of B's and C's for males as a function of affirmation condition [$\chi^2(1, N=202)=0.02$, $p=0.88$].

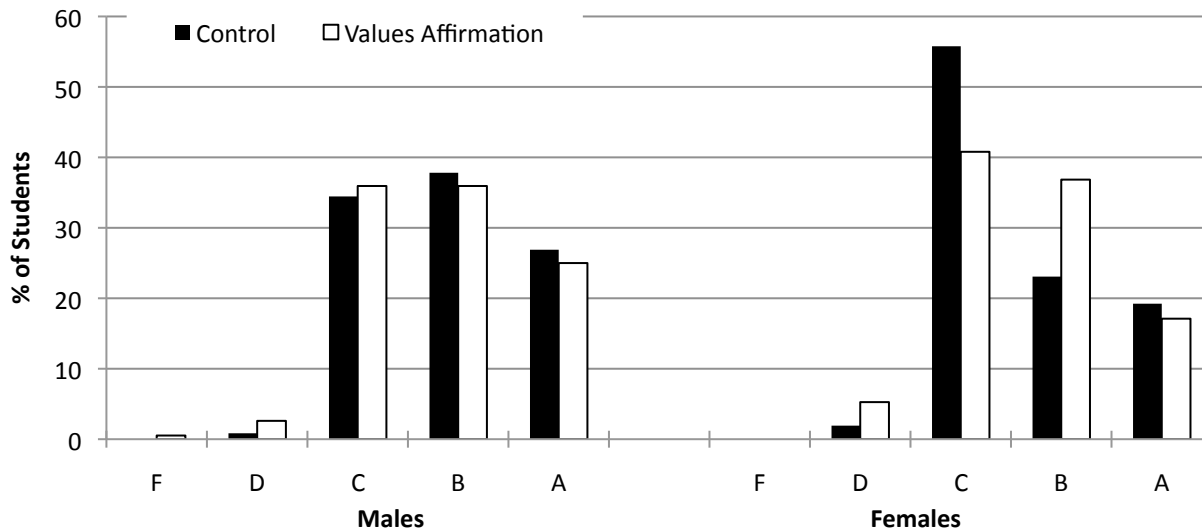


Figure 5. Percentage of students receiving each letter grade (A, B, C, D, and F, combining letter grades with pluses and minuses) as a function of gender and affirmation condition. The percentage was calculated separately for each condition within gender. The distribution of grades for males is not significantly different by condition, but the distribution of grades for females is statistically different for females in the control and affirmation conditions.

C. Force and Motion Conceptual Evaluation

The FMCE was administered twice as part of the course, once at the beginning of the semester and once at the end, to assess learning of physics concepts over the semester. Although students completed the first affirmation or control exercise immediately before taking the FMCE pre-test, there was no significant effect of the intervention on the FMCE pre-test score. There was an overall gender gap on the FMCE pre-test across the entire sample, with males ($M=39.9\%$) outscoring females ($M=25.0\%$) [$\beta=-0.27$, $t(304)=-4.85$, $p<0.01$]. The gender gaps in the affirmation and control conditions were not significantly different, as evidenced by the insignificant *gender* \times *condition* interaction [$\beta=0.04$, $t(304)=0.73$, $p=0.47$].

Though there was no significant effect of the intervention on the FMCE pre-test, there was an effect on the post-test. Students' FMCE post-test scores were regressed on gender, condition, stereotype endorsement, and FMCE pre-test score, as well as the interactions of interest. The regression results are presented in Table 3.

Table 3. Multiple regression coefficient estimates and significance levels. The dependent variable is the FMCE post-course score. The independent variables significantly predicted students' FMCE post-test scores [$F(1,296)=15.69, p<0.01$]. This model accounts for 37% of the variance in mean overall exam score ($R^2=0.37$).

Variables	β	$t(296)$	Significance
Gender	-0.06	-1.05	0.30
Condition	0.10	1.93	0.06
Stereotype Endorsement	-0.10	-1.67	0.10
FMCE Pre-test Score	0.66	8.80	<0.01
Gender \times Condition	0.12	2.13	0.03
Gender \times Stereotype Endorsement	0.01	0.19	0.85
Gender \times FMCE Pre-test Score	0.17	2.40	0.02
Condition \times Stereotype Endorsement	0.15	2.69	<0.01
Condition \times FMCE Pre-test Score	0.03	0.53	0.60
Stereotype Endorsement \times FMCE Pre-test Score	0.07	1.42	0.16
Gender \times Condition \times Stereotype Endorsement	0.15	2.45	0.02

There was an overall gender gap across the entire sample on the FMCE post-test, with males ($M=73.4\%$) demonstrating better conceptual mastery at the end of the semester than females ($M=60.4\%$) [$\beta=-0.06, t(296)=-1.05, p=0.30$]. This gender effect, however, was moderated by the affirmation condition, as reflected in the predicted *gender \times condition* interaction [$\beta=0.12, t(296)=2.13, p=0.03$]. As shown in Figure 6, the gender gap in the control condition was significant ($d=0.46$) [$F(1,296)=6.23, p=0.01$], whereas the gender gap in the affirmation condition was not ($d=-0.12$) [$F(1,296)=0.96, p=0.33$]. Unlike for the exams, for the FMCE, the reduction of the gender gap in the affirmation condition was due almost entirely to females' increased score in the affirmation condition. Specifically, females in the affirmation condition had significantly higher FMCE scores than females in the control condition [$F(1,296)=7.71, p<0.01$], whereas the scores of males in the two conditions did not differ [$F(1,296)=0.08, p=0.78$]. Thus, even though there was some unexpected tendency for males in the affirmation condition to perform worse than males in the control condition for the two highly

correlated measures of the composite exam score and the final course score, there was no such effect for the end-of-semester FMCE data.

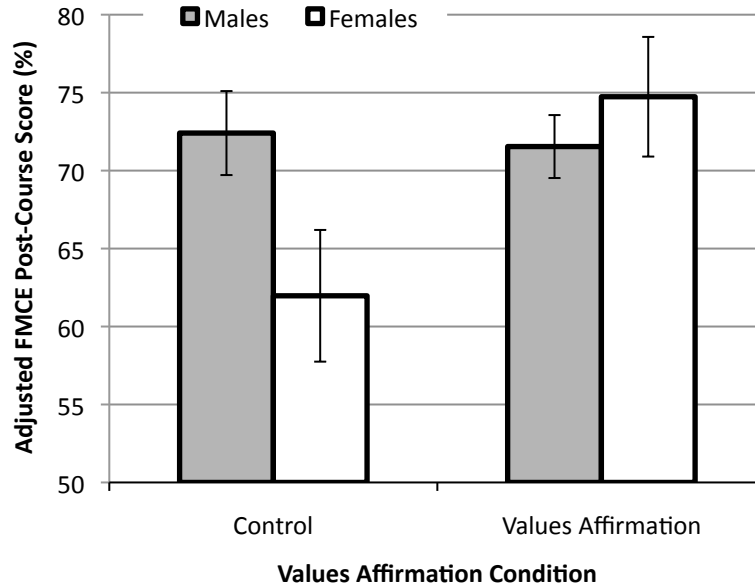


Figure 6. Student performance on the FMCE post-test as a function of gender and condition. FMCE post-test scores are adjusted based on FMCE pre-test scores. Error bars represent the standard error on the mean. The gender gap in the control condition is statistically significant ($p < 0.01$), but the gender gap in the affirmation condition is not statistically significant ($p = 0.33$).

Finally, the effects of values affirmation on the FMCE post-test performance of females were once again moderated by the level of gender stereotype endorsement, as seen in Figure 7. The predicted *gender* \times *condition* \times *stereotype endorsement* interaction [$\beta = 0.15$, $t(296) = 2.45$, $p = 0.02$] was significant. Decomposing the three-way interaction, we found no significant relationship between stereotype endorsement and FMCE scores for males in either the affirmation or control condition [$\beta = -0.10$, $t(296) = -1.41$, $p = 0.16$, and $\beta = -0.13$, $t(296) = -1.33$, $p = 0.18$, respectively]. For females, however, there was a significant negative relationship between stereotype endorsement and FMCE scores only in the control condition [$\beta = -0.39$, $t(296) = -2.55$, $p = 0.01$]. By contrast, for females in the affirmation condition, the relationship was not significant [$\beta = 0.22$, $t(296) = 1.54$, $p = 0.13$]. Among the females with higher stereotype

endorsement (0.75 SDs above the mean), the end-of-semester FMCE scores were significantly higher in the affirmation condition than in the control condition [$t(115)=3.01, p<0.01$].

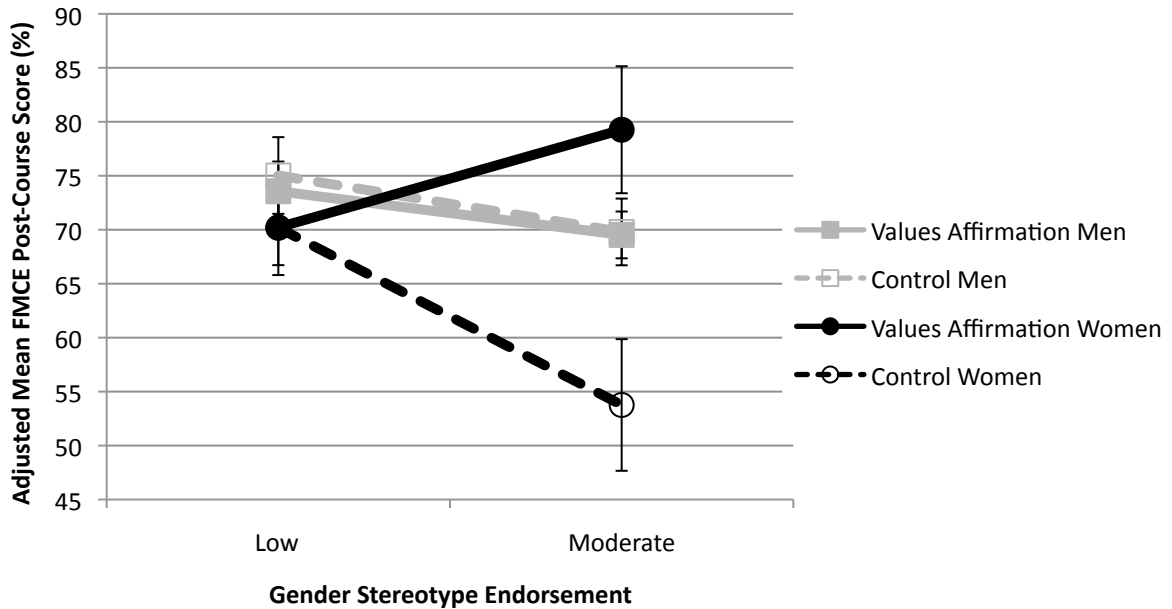


Figure 7. Student performance on the FMCE post-test as a function of gender, condition, and the level of stereotype endorsement. The FMCE post-test score is adjusted based on the FMCE pre-test score. The error bars represent the standard error on the mean. Moderate and low gender stereotype endorsement represent ± 0.75 SD of the grand mean of stereotype endorsement. The relationship between mean exam score and stereotype endorsement is only significant for females in the control group. Among females who moderately endorsed the gender stereotype, females in the control group had significantly lower mean exam scores than females in the affirmation group.

D. Historical Comparison

Because the FMCE is a standardized instrument (unlike in-class exams, which vary from semester to semester), we can compare student performance this semester to past semesters on the same set of questions. Additionally, the instructor who taught during the semester of the study taught the same course the previous year, so not only can we compare to prior offerings of the course, but also to a prior semester with the same instructor. Figure 8 shows the male and female average FMCE post-test score (adjusted for FMCE pre-test score) for the control and affirmation conditions from the current semester, for students who took the class in a previous semester with the same instructor, and for students who took Physics 1 in all prior semesters in

which FMCE data were collected. First, we see that the average FMCE post-test scores of this instructor are significantly higher than the average FMCE post-test scores over all semesters. Males in both the affirmation and control conditions do not have FMCE post-test scores that are significantly different from males' scores last time this instructor taught. While females in the affirmation group performed better than females did the last time the instructor taught (as we expected), what is unexpected is that females in the control group performed worse than females in this instructor's last class. The control group females' lower performance results in a significantly larger gender gap among the control group students compared to the last time this instructor taught. Since we didn't expect any impact on student performance of the control writing exercise, it is curious that the gender gap in the control group is so much larger than in prior semesters (though there have been examples of individual semesters, not taught by this instructor, when the gender gap was as large).

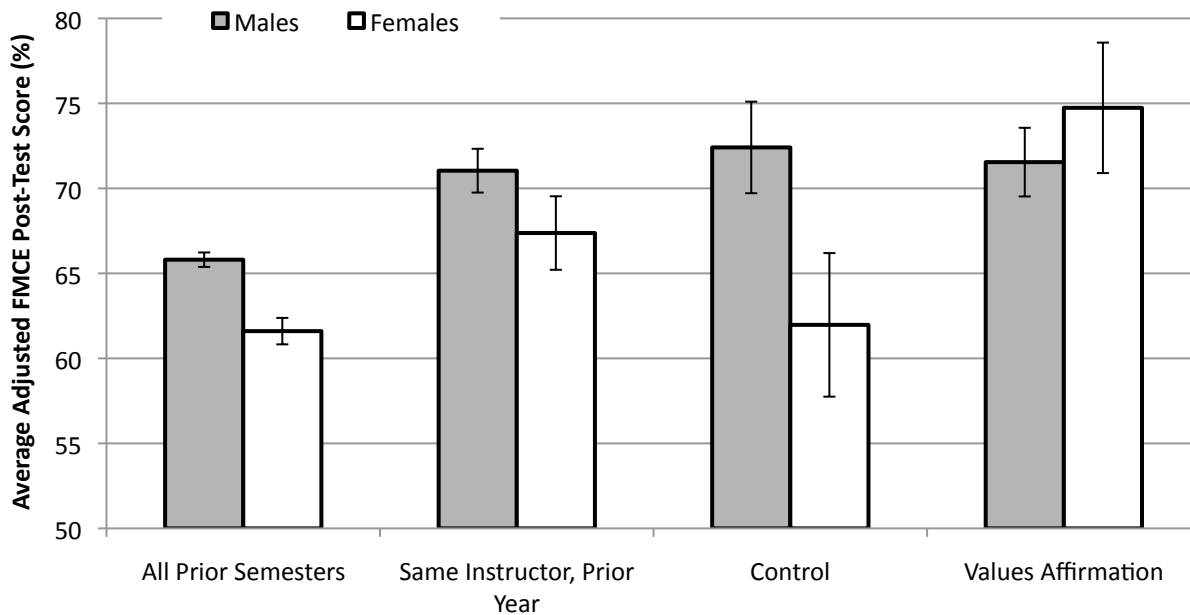


Figure 8. Male and female average FMCE post-test performance for all semesters of Physics 1 in which pre- and post-FMCE data were collected, for students in a prior semester in which the same instructor taught, and the control and affirmation groups from the current study.

E. Other Course Components

Though we were mainly interested in the intervention's effect on students' exam and conceptual survey scores, we also examined the impact of the intervention on students' homework, participation, and clicker scores. Recall from prior chapters that females often outscore males on these aspects of the course. As in Figure 9, that is the case in this semester for students in the control group. For all four measures, females in the control group have significantly higher scores than males in the control group [F 's(1,392)>5.65, p 's<0.02] (details of the regression analyses for each measure can be found in Appendix E). For students in the affirmation condition, however, females and males do not have significantly different scores on any of the measures [F 's(1,392)<2.65, p 's>0.10]. Despite the differences in gender gaps between the affirmation and control conditions, the intervention had a statistically significant effect only on students' CAPA homework scores and their clicker scores, as evidenced by the significance of the *gender* × *condition* interaction [β =-0.15, $t(392)=-2.60$, $p=0.01$ and β =-0.12, $t(392)=-2.19$, $p=0.03$, respectively]. When examining the simple effects of the intervention on males and females separately, the trends are opposite of what we observe for exams and the FMCE. Females in the affirmation condition have *lower* scores than females in the control condition, and males in the affirmation condition have *higher* scores than males in the control condition. These effects are only significant for females' CAPA homework scores [$F(1,392)=5.15$, $p=0.02$] and males' clicker scores [$F(1,392)=5.21$, $p=0.02$], but the trends are the same across all four measures.

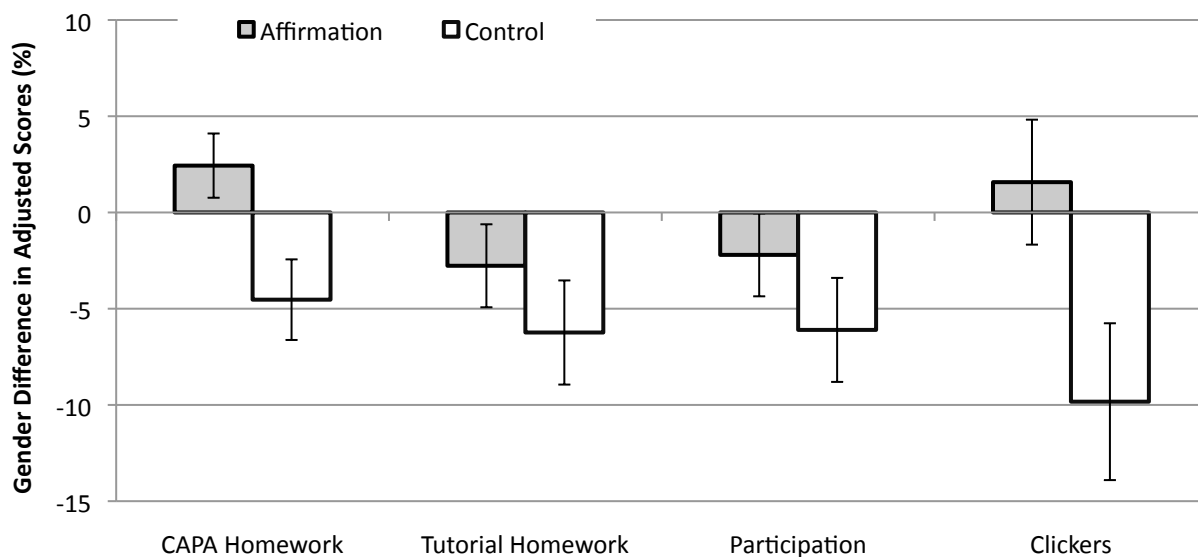


Figure 9. Gender differences on homework, participation and clickers as a function of condition. All scores are adjusted based on students' prior math score. Error bars represent the standard error on the mean. Females in the control group outscore males in the control group on all four measures, while the gender differences in the affirmation condition are not statistically significant for any of the four measures.

F. Replication Study

In the semester following the original study we conducted a replication study. The course was the same, Physics 1, but the instructor was different. Though the curriculum and pedagogy of the class (Peer Instruction [1] and *Tutorials* [2] were used) was the same in the replication semester, the instructor was less experienced using interactive engagement techniques. The implementation of the interactive engagement techniques differed from the instructor in the original study, and the exams were different. We attempted to replicate the original study design as closely as possible, though there were some problems with the implementation of the first writing exercise [54]. Additionally, we modified the pre-course survey in the replication study. The original survey contained not only the stereotype endorsement item that we discussed above, but also an additional nine items related to students' perception that they and others endorsed the stereotype that men are better at physics than women. For the replication study, we removed

eight of the ten stereotype endorsement items (keeping only the item of interest and its corresponding statement, *According to my own personal beliefs, I expect women to do better than men in physics*). This was done to minimize negative student reactions to the survey and to lessen the chance that we invoked identity threat among students through the survey.

The conditions for including students in the replication sample of students were the same as in the original study. Students were included if: a) they consented to be in the study and were over 18 years of age, b) they completed both writing exercises correctly, c) they took the final exam, d) we had data on either their SAT- or ACT-Math score, e) they responded to the stereotype endorsement item, and f) they were not an outlier based on multivariate outlier analyses [41]. Unfortunately, fewer students met all of these criteria in the replication semester. Fewer students completed the first writing exercise compared to the prior semester (541 versus 591), and then fewer of those students completed the second writing exercise (69% versus 77%). The final study sample was made up of 127 males and 60 females in the affirmation condition and 85 males and 19 females in the control condition. Of the 291 students in the final study sample, only 217 (75%) took both the FMCE pre- and post-test and were included in the FMCE sample. The numbers of students in each sample broken out by gender and condition are presented in Table 4. Though the small sample sizes make it difficult to make any statistical claims about the effectiveness of the self-affirmation intervention, we present the results regardless, and focus on the trends of the results to see if they are consistent with the original study.

Table 4. Numbers of males and females in the affirmation and control conditions for each of the samples used in the replication study.

	Males	Females
Final Study Sample (N=291)		
Affirmation	127	60
Control	85	19
FMCE Sample (N=217)		
Affirmation	94	47
Control	56	17

We repeated the statistical analysis used in the original study, using regression analysis to test for the significance of the *gender × condition* and *gender × condition × stereotype endorsement* interactions in exam and FMCE post-test scores. The regression analysis results are presented in Table 5 for the mean overall exam scores and Table 6 for the FMCE post-test scores. For both outcome measures, the *gender × condition* interaction was not significant [$\beta=-0.04$, $t(279)=-0.61$, $p=0.54$ for the exams, and $\beta=0.03$, $t(205)=0.48$, $p=0.63$ for the FMCE]. This suggests that there was not a statistically significant difference in the gender gaps between the two conditions. The *gender × condition × stereotype endorsement* interaction was also not significant [$\beta=-0.03$, $t(279)=-0.48$, $p=0.64$ for the exams, and $\beta=0.04$, $t(205)=0.50$, $p=0.62$ for the FMCE]. We do not see the same results as we did in the original study, that is, there is no evidence to suggest that the intervention had any impact on student performance.

Table 5. Multiple regression coefficient estimates and significance levels. The dependent variable is the mean overall exam score. The independent variables significantly predicted students' mean overall exam scores [$F(1,279)=12.97, p<0.01$]. This model accounts for 34% of the variance in mean overall exam score ($R^2=0.34$).

Variables	β	$t(279)$	Significance
Gender	-0.15	-2.54	0.01
Condition	-0.07	-1.11	0.27
Stereotype Endorsement	-0.04	-0.55	0.58
Math Score	0.54	9.38	<0.01
Gender \times Condition	-0.04	-0.61	0.54
Gender \times Stereotype Endorsement	0.07	1.00	0.32
Gender \times Math Score	0.001	0.01	0.99
Condition \times Stereotype Endorsement	-0.02	-0.27	0.79
Condition \times Math Score	-0.02	-0.42	0.67
Stereotype Endorsement \times Math Score	0.01	0.20	0.84
Gender \times Condition \times Stereotype Endorsement	-0.03	-0.48	0.64

Table 6. Multiple regression coefficient estimates and significance levels. The dependent variable is the FMCE post-course score. The independent variables significantly predicted students' FMCE post-test scores [$F(1,205)=13.38, p<0.01$]. This model accounts for 42% of the variance in mean overall exam score ($R^2=0.42$).

Variables	β	$t(205)$	Significance
Gender	0.04	0.60	0.55
Condition	-0.04	-0.61	0.55
Stereotype Endorsement	-0.01	-0.15	0.88
FMCE Pre-test Score	0.74	8.04	<0.001
Gender \times Condition	0.03	0.48	0.63
Gender \times Stereotype Endorsement	0.03	0.40	0.69
Gender \times FMCE Pre-test Score	0.18	2.15	0.03
Condition \times Stereotype Endorsement	-0.03	-0.38	0.71
Condition \times FMCE Pre-test Score	0.07	1.17	0.24
Stereotype Endorsement \times FMCE Pre-test Score	0.05	0.93	0.35
Gender \times Condition \times Stereotype Endorsement	0.04	0.50	0.62

As mentioned above, the lack of statistical significance in the results is not surprising given the small sample sizes. We therefore examine the trends in the data to see if they are the same as in the original study. Figure 10 presents the male and female average exam and FMCE post-test scores by condition, adjusted for prior math score or FMCE pre-test score. We did not

find that the trends in the replication study were the same as in the original study. Looking first at the mean overall exam scores, there is about a 4% gender gap in exam scores for students in the control condition. Unexpectedly, the gender gap is *larger*, about 6%, among students in the affirmation condition, and is statistically significant [$F(1,279)=12.88, p<0.01$]. It appears that the affirmation intervention may have hurt female performance on exams in this semester.

Looking at the FMCE post-test scores, again, we do not see trends similar to those observed in the original study. We find that after controlling for FMCE pre-test score, there is effectively no gender gap in FMCE post-test scores among students in the control condition. In the affirmation condition, females outscore males by about 4%, similar to what was observed in the affirmation condition for FMCE post-test scores in the previous study. Comparing students within each gender, females' scores are virtually the same in the two conditions, while males in the affirmation condition had lower FMCE post-test scores than males in the control condition.

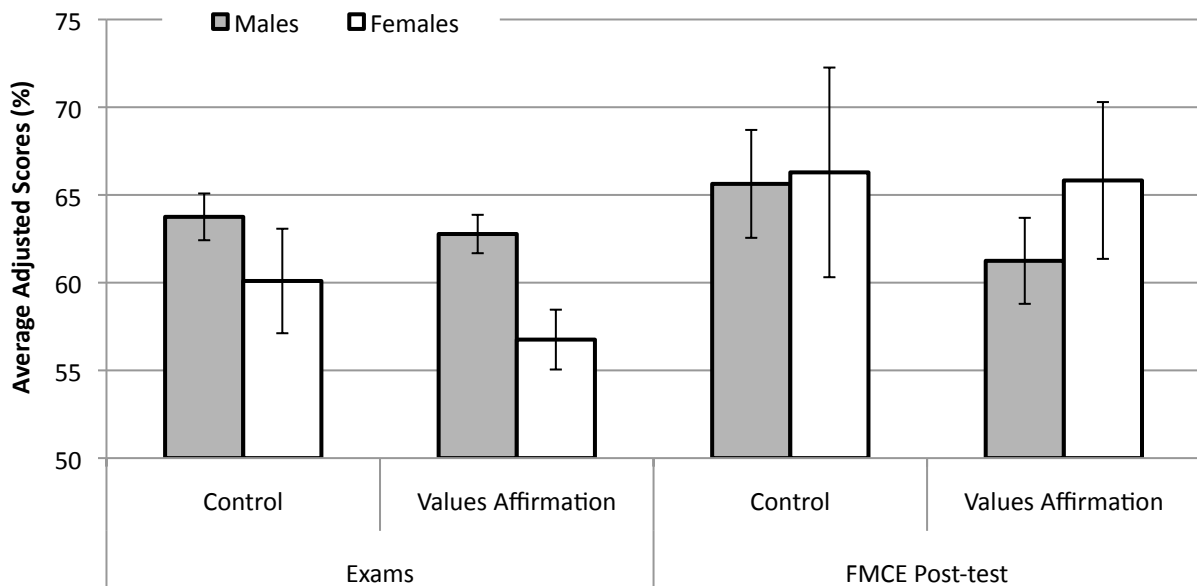


Figure 10. Average adjusted overall mean exam and FMCE post-test performance as a function of gender and condition. Exam scores are adjusted based on prior math scores, and the FMCE post-test scores are adjusted based on FMCE pre-test scores. Error bars represent the standard error on the mean. Females in the affirmation condition had significantly lower exam scores than males in the affirmation condition [$F(1,279)=12.88, p<0.01$]. All other gender differences are not statistically significant [F 's(1,279)<1.95, p 's>0.15].

Given the intervention implementation issues, the small sample sizes, and the inconsistent results, it is difficult to draw any conclusions from the replication study. The FMCE results are inconclusive and the exam results suggest that the affirmation intervention was not successful in reducing or eliminating the gender gap, and may have made it worse. If anything, maybe we can conclude, as others have [37], that the affirmation intervention is not a guaranteed solution and needs to be implemented in concert with practices that promote student learning and in a classroom environment that is supportive of all students.

V. Discussion and Conclusions

Overall, these results suggest that values affirmation is a promising intervention that can help reduce the gender achievement gap in physics. Although the intervention was brief and did not directly concern the course material, it nonetheless provided a meaningful boost for females—especially females who tended to endorse the gender stereotype—on two objective measures, in-class multiple-choice exams and the FMCE post-test. After controlling for prior background (prior math score or FMCE pre-test scores), the affirmation closed the “residual” gender gap on in-class exam scores by approximately 61% and entirely eliminated the gap on the FMCE. Although further efforts must aim to close the gap in prior preparation, the current results are promising in demonstrating that a brief psychological intervention can help close the "residual" gap. Further, the introductory physics class we investigated was intended for STEM majors who have had success in STEM-related subjects before college and are motivated to do well in the course. Our results, therefore, demonstrate that, even among females relatively identified and accomplished in science, a substantial gender gap exists, females’ performance is negatively related to stereotype endorsement, and gender differences can be reduced with a values-affirmation intervention.

Hypothesized decreases in identity threat that occur following affirmation could benefit females by improving either test performance, actual learning of new concepts, or both. We cannot differentiate between the contribution of solely performance-related and solely learning-related benefits on exam scores and final course scores. However, the results on the FMCE post-test provide novel evidence that identity threat and affirmation affect the learning of scientific concepts, not just performance. Because evaluation apprehension was low for the FMCE (students took this test in recitation both times and were told that their performance on the FMCE would not affect their grade) and because performance was assessed on identical standardized items across the semester, FMCE performance provides a better indication of actual learning effects. The improvement among females following the affirmation on the FMCE post-test (with the FMCE pre-test controlled for) is, therefore, promising evidence that affirmation can produce beneficial effects through the facilitation of better learning. Additionally, the absence of any affirmation effect on the FMCE pre-test scores suggests that the FMCE post-test effects were not solely due to better performance, but, rather, that values affirmation likely had benefits on the actual learning of new concepts over the semester.

There was no significant difference in the performance of males in the control and affirmation conditions on the FMCE post-test score. However, we obtained an unexpected effect of affirmation on the performance of males on exam scores and course scores. The pattern was such that, although affirmation improved performance for females relative to the control condition, it decreased performance for males (this negative effect for males was significant for exam scores and marginal for course scores). This pattern was not predicted and was not obtained in the original affirmation field experiments [37]. Because this pattern was not consistently observed either in the past research or across all measures in the present study (e.g.,

FMCE post-test, letter grade distribution), it should be regarded tentatively. At the same time, though not predicted here, negative effects of affirmation have sometimes been observed in the affirmation literature [39]. More research needs to be conducted to specify when and why such negative effects may occur [55].

Despite the lower exam scores for males in the affirmation condition, it is important to emphasize that such effects were not observed on the FMCE post-test score and, perhaps more important, that the reduction in the gender gap associated with affirmation does not simply reflect a negative effect of affirmation on males. As noted above, the affirmation significantly improved females' performance relative to those in the control group across all main outcome measures (exam, FMCE, and course scores). Such consistent results found for females indicate that, on the whole, affirmation closed the gender gap by directly benefitting females.

While the main findings of this study were as we predicted, there were some aspects of the results that we find curious. While not the focus of our study, we examined the impact of the affirmation intervention on students' homework, participation, and clicker scores. The results were not consistent with the exam and FMCE results, that is, females in the affirmation condition had *lower* homework, participation, and clicker scores than females in the control condition. These results are not well understood. Prior research on identity threat and self-affirmation is primarily concerned with test performance, as identity threat effects are strongest for more difficult tasks [10]. Further investigation is needed to understand the effect of identity threat and self-affirmation on low-stakes, untimed, effort-based assessments. A second unexpected finding was the reduced FMCE performance of females in the control group compared to females in prior semesters (Figure 8). We did not expect the control writing exercises to have any impact on student performance. One factor that may have played a role was the student attitude survey that

was given in the second week of the course. As mentioned above, there were a total of ten items on the survey that asked students about their own and their perception of others' endorsement of the stereotype that men are better than women at physics. Student reactions to the survey were more negative than student reactions to similar pre-course surveys (that did not contain any references to gender) in prior semesters. These items may have heightened students' awareness of gender issues in the physics classroom, and perhaps increased the identity threat that we were attempting to alleviate (leading to an increased gender gap among control students compared to prior semesters).

Although previous attempts to reduce the gender achievement gap in physics have focused mostly on instructional methods [5,6,7,8], the current results highlight the importance of social-psychological factors. One virtue of the affirmation is that it can be combined with instructional approaches that show promise in closing the gender gap [5], such as the interactive engagement approaches used in the present course [1,2]. However, there is no reason to think that the effects of affirmation are confined to situations in which such instructional supports are already in place, given that the intervention was successful in reducing racial achievement differences among middle-school students in traditional classrooms [37,56]. Of course, even in that study, there were structural opportunities for learning in the form of a solid curriculum and qualified teachers; without such basic support, the efficacy of any psychological intervention would be limited [37]. Results from our replication study may support this claim.

Finally, the benefits of the affirmation may be long lasting [11,56] and could persist beyond the present course. The experience of lower identity threat, coupled with better mastery in a challenging science course, may encourage affirmed females to take other STEM courses and to pursue further education and even a career in STEM disciplines. More generally, the

cumulative consequences of early performance—small differences at an early stage can get magnified over time—help explain how relatively brief interventions, when given early in a threatening environment, can have long-term effects. This snowballing effect may be particularly important in science, where subsequent learning builds on an earlier foundation of knowledge, making it increasingly difficult to catch up and enter a discipline later. Therefore, it may be important to intervene in gateway courses like introductory physics. These courses are required for STEM majors, and performance in them can set long-term academic and career trajectories. Reducing the gender gap at gateways could not only benefit females’ performance in the short term but also encourage them to choose and persist in a scientific major and career path in STEM disciplines.

Peer-reviewed publications based on this work

- A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen & T. A. Ito, Reducing the gender achievement gap in college science: A classroom study of values affirmation, *Science* **330** 1234 (2010).
- L. E. Kost-Smith, S. J. Pollock, N. D. Finkelstein, G. L. Cohen, T. A. Ito and A. Miyake, “Gender Differences in Physics 1: The Impact of a Self-Affirmation Intervention” in *Proceedings of the 2010 Physics Education Research Conference*, edited by M. Sabella, C. Singh and S. Rebello (AIP Conference Proceedings, New York, 2010) p. 197.

References (Chapter 6)

- [1] E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- [2] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- [3] P. Heller, R. Keith and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *American Journal of Physics* **60**, 627 (1992).
- [4] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *American Journal of Physics* **60**, 637 (1992).
- [5] M. Lorenzo, C. Crouch and E. Mazur, Reducing the gender gap in the physics classroom, *American Journal of Physics* **74**, 118 (2006).
- [6] E. Brewster, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez and P. Pamelá, Toward equity through participation in Modeling Instruction in introductory university physics, *Physical Review Special Topics - Physics Education Research* **6**, 010106 (2010).
- [7] P. Laws, P. Rosborough and F. Poodry, Women's responses to an activity-based introductory physics program, *American Journal of Physics* **67**, S32 (1999).
- [8] S. Brahmia and E. Etkina, Switching students on to science, *Journal of College Science Teaching* **31**(3), 183 (2001).
- [9] M. Schneider, Encouragement of women physics majors at Grinnell college: A case study, *Physics Teacher* **39**, 280 (2001).
- [10] C. M. Steele, S. J. Spencer and J. Aronson, "Contending with group image: The psychology of stereotype and social identity threat," in *Advances in Experimental Social Psychology*, edited by M. P. Zanna (Academic Press, New York, 2002), p. 379.
- [11] G. L. Cohen and J. Garcia, Identity, belonging, and achievement: A model, interventions, and implications, *Current Directions in Psychological Science* **17**(6), 365 (2008).
- [12] G. M. Walton and G. L. Cohen, A question of belonging: Race, social fit, and achievement, *Journal of Personality and Social Psychology* **92**(1), 82 (2007).
- [13] G. L. Cohen and J. Garcia, "I am us": Negative stereotypes as collective threats, *Journal of Personality and Social Psychology* **89**, 566 (2005).
- [14] C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance, *American Psychologist* **52**, 613 (1997).

- [15] S. J. Spencer, C. M. Steele, and D. M. Quinn, Stereotype threat and women's math performance, *Journal of Experimental Social Psychology* **35**, 4 (1999).
- [16] S. L. Beilock, R. J. Rydell, and A. R. McConnell, Stereotype threat and working memory: Mechanisms, alleviation and spillover, *Journal of Experimental Psychology: General* **136**(2), 256 (2007).
- [17] C. M. Steele and J. Aronson, Stereotype threat and the intellectual test performance of African Americans, *Journal of Personality and Social Psychology* **69**, 797 (1995).
- [18] J. L. Brown and R. A. Josephs, A burden of proof: Stereotype relevance and gender differences in math performance, *Journal of Personality and Social Psychology* **76**, 246 (1999).
- [19] M. Inzlicht and T. Ben-Zeev, A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males, *Psychological Science* **11**, 365 (2000).
- [20] M. Shih, T. L. Pittinsky and N. Ambady, Stereotype susceptibility: Identity salience and shifts in quantitative performance, *Psychological Science* **10**, 80 (1999).
- [21] C. Good, J. Aronson and J. A. Harder, Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses, *Journal of Applied Developmental Psychology* **29**, 17 (2008).
- [22] N. Ambady, M. Shih, A. Kim and T. L. Pittinsky, Stereotype susceptibility in children: Effects of identity activation on quantitative performance, *Psychological Science* **12**, 385 (2001).
- [23] B. Cole, K. Matheson and H. Anisman, The moderating role of ethnic identity and social support on relations between well-being and academic performance, *Journal of Applied Social Psychology* **37**, 592 (2007).
- [24] J. Aronson, M. Lustina, C. Good, K. Keough, C. M. Steele and J. Brown, When white men can't do math: Necessary and sufficient factors in stereotype threat, *Journal of Experimental Social Psychology* **35**, 29 (1999).
- [25] J. Stone, C. I. Lynch, M. Sjomeling and J. M. Darley, Stereotype threat effects on Black and White athletic performance, *Journal of Personality and Social Psychology* **77**, 1213 (1999).
- [26] J. Stone and C. McWhinnie, Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes, *Journal of Experimental Social Psychology* **44**, 445 (2008).

- [27] L. J. Kray, A. D. Galinsky and L. Thompson, Reversing the gender gap in negotiations: An exploration of stereotype regeneration, *Organizational Behavior and Human Decision Processes* **87**, 386 (2002).
- [28] B. Levy, Improving memory in old age through implicit self stereotyping, *Journal of Personality and Social Psychology* **71**, 1092 (1996).
- [29] N. C. Yeung and C. von Hippel, Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers, *Accident Analysis and Prevention* **40**, 667 (2008).
- [30] T. Schmader, M. Johns and M. Barquissau, The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain, *Sex Roles: A Journal of Research* **50**, 835 (2004).
- [31] J. Aronson and M. Inzlicht, The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African American college students, *Psychological Science* **15**, 829 (2004).
- [32] J. Aronson, C. B. Fried and C. Good, Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence, *Journal of Experimental Social Psychology* **38**, 113 (2002).
- [33] C. Good, J. Aronson and M. Inzlicht, Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat, *Applied Developmental Psychology* **24**, 645 (2003).
- [34] T. Schmader, Gender identification moderates stereotype threat effects on women's math performance, *Journal of Experimental Social Psychology* **38**, 194 (2002).
- [35] B. Hufnagel, G. L. Deming, J. M. Landato and A. K. Hodari, The effect of stereotype threat on undergraduates in an introductory astronomy class, *Journal of Women and Minorities in Science and Engineering* **10**, 89 (2004).
- [36] A. Martens, M. Johns, J. Greenberg and J. Schimel, Combating stereotype threat: The effect of self-affirmation on women's intellectual performance, *Journal of Experimental Social Psychology* **42**, 236 (2006).
- [37] G. L. Cohen, J. Garcia, N. Apfel and A. Master, Reducing the racial achievement gap: A social-psychological intervention, *Science* **313**, 1307 (2006).
- [38] C. M. Steele, "The psychology of self-affirmation: Sustaining the integrity of the self" in *Advances in Experimental Social Psychology*, edited by L. Berkowitz (Academic Press, New York, 1988), p. 261.

- [39] D. K. Sherman and G. L. Cohen, “The psychology of self-defense: Self-affirmation theory” in *Advances in Experimental Social Psychology*, edited by M. P. Zanna (Academic Press, San Diego, CA, 2006), p. 183.
- [40] V. Otero, N. D. Finkelstein, R. McCray and S. Pollock, Who is responsible for preparing science teachers? *Science* **313**, 445 (2006).
- [41] C. M. Judd and G. H. McClelland, *Data Analysis: A Model Comparison Approach* (Harcourt Brace Jovanovich, New York, 1989).
- [42] The Mahalanobis distance gives an estimate of how different a set of measurements for one individual is from the average set of measurements for some population. It can be calculated using a statistical analysis program, such as SPSS.
- [43] The twelve background measures included: high school GPA, SAT-Verbal score, SAT-Math score, SAT-Total score, ACT-Math score, ACT-Reading score, ACT-English score, ACT-Science Reasoning score, ACT-Total score, math score (a normalized average of the SAT- and ACT-Math scores), years of high school physics, and years of high school calculus.
- [44] G. L. Cohen, J. Aronson and C. M. Steele, When beliefs yield to evidence: Reducing biased evaluation by affirming the self, *Personality and Social Psychology Bulletin* **26**, 1151 (2000).
- [45] The values affirmation exercise that we used in this study was used in prior studies with middle-school age children. The language of the exercise was modified for college age adults. Additionally, the instructor of the course was sensitive to students’ comfort in writing about personal values in his course. To avoid making students uncomfortable, some of the values were modified to make them less sensitive, for instance “religious values” was changed to “spiritual or religious values” and “politics” was changed to “government and politics”. Also, at the instructor’s request, “learning and gaining knowledge” was added to the list of possible values.
- [46] W. K. Adams, K. K. Perkins, N. Podolefsky, M. Dubson, N. D. Finkelstein and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: the Colorado Learning Attitudes about Science Survey, *Physical Review Special Topics - Physics Education Research* **2**, 010101 (2006).
- [47] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton’s laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *American Journal of Physics* **66**, 338 (2006).
- [48] CAPA, <http://www.lon.capa.org/>.

- [49] L. S. Aiken and S. G. West, *Multiple Regression: Testing and Interpreting Interactions* (Sage, Thousand Oaks, CA, 1991).
- [50] T. Schmader and M. Johns, Converging evidence that stereotype threat reduces working memory capacity, *Journal of Personality and Social Psychology* **85**, 440. (2003).
- [51] Z. Hazari, R. H. Tai and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, *Science Education* **91**, 847 (2007).
- [52] V. Y. Yzerbyt, D. Muller and C. M. Judd. Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions, *Journal of Experimental Social Psychology* **40**, 424 (2004).
- [53] Here, effect size is calculated as $ES = (\langle S \rangle_{males} - \langle S \rangle_{females}) / SD_{pooled}$, where $SD_{pooled} = [(N_{males} - 1) * SD_{males}^2 + (N_{females} - 1) * SD_{females}^2] / (N_{males} + N_{females})^{1/2}$.
- [54] Though the TAs were given the same script to follow in the replication semester, study personnel who observed them during the first recitation found that several of the TAs did not follow the scripts. There were also issues with the order of activities; at least one TA tried to give students the FMCE before doing the writing exercise, but the study personnel interrupted and told him to do the writing exercise first. Unlike in the original study, the study personnel did not remain discreet, as they had to intervene during some recitation sections. It is not clear what, if any, impact these issues had on the effectiveness of the intervention.
- [55] In the present case, there are possible reasons why affirmation could be counterproductive for a person belonging to a group free of pervasive threats tied to their social identity. For instance, affirmation might divert their attention away from the performance domain, remind them of alternative domains where they could invest their effort, or make their sense of self-integrity less dependent on performing well. These are admittedly speculative explanations for a finding that should, as noted, be regarded tentatively.
- [56] G. L. Cohen, J. Garcia, V. Purdie-Vaughns, N. Apfel and P. Brzustoski, Recursive processes in self-affirmation: Intervening to close the minority achievement gap, *Science* **324**, 400 (2009).

CHAPTER 7: CONCLUSIONS

In this final chapter, we take the opportunity to summarize the findings across all of the research studies, discuss the implications of these results for the introductory physics courses, and recommend directions for future research on the underrepresentation of females in physics.

I. Summary of Findings

The goal of our work is to understand differences in the participation and performance of males and females in physics. We have focused on the introductory physics courses at the University of Colorado, as success in these courses is required for further study in physics as well as other science, engineering, and mathematics majors. Our methodology was primarily quantitative, employing an epidemiological approach and statistical analysis techniques to characterize gender differences and model student performance in these courses. Data were gathered from university records, course instructors, and students through surveys developed by the author. In addition to documenting and modeling gender disparities in the introductory physics courses, we also sought to address the disparities in performance by implementing and testing the effectiveness of a social psychological intervention. Here, we summarize the major findings of our studies, organized by the four research questions that we set out to answer.

1. *On what measures, and to what extent, do we observe differences by gender in the introductory physics courses? For example, are there gender differences in conceptual learning, course grades and their components, attitudes and beliefs, and prior knowledge and preparation?*

We identified gender differences in several aspects of the introductory physics courses, including students' background and preparation coming into the physics courses, their performance as they worked through the introductory physics courses, and whether students continue in the introductory physics sequence from Physics 1 to Physics 2. In most cases, the gender differences favor males. Males are more prepared for the introductory physics courses than females; in Chapters 2 and 3 (for both Physics 1 and Physics 2) we found that males are more likely to have taken a high school physics course, they have higher standardized math test scores, and they have higher conceptual physics pre-test scores than females. In addition to their physics and mathematics preparation, males also have more favorable attitudes and beliefs about physics and about learning physics coming into the introductory courses than females.

Male students outperform female students in the introductory physics courses, though not universally. Post-course conceptual survey scores of males are higher than the scores of females by 5% to 18% in Physics 1 and by 4% to 10% in Physics 2. The average effect sizes of these differences are 0.46 for Physics 1 and 0.44 in Physics 2, indicating that males are outperforming females on a post-course measure of conceptual physics understanding by almost half of a standard deviation. These post-course gender differences existed, even in semesters where several interactive engagement strategies [1,2,3] were used. In addition to outscoring females on a diagnostic conceptual survey, males also score higher on in-class exams than females by 5% in Physics 1 and 4% in Physics 2. While males outscore females on high-pressure, time-sensitive, individual tests, females outscore males on the untimed, potentially collaborative homework

assignments and on effort-based participation. These two effects generally offset one another, such that the course grades of males and females in most individual semesters are not significantly different, but because exams are weighted slightly more than homework and participation, there is a significant gender difference in course grades when averaging over all semesters of Physics 1, favoring males.

In addition to gender differences in performance that we observed throughout the introductory physics courses, there are also gender differences in how students' attitudes and beliefs shift over the course of Physics 1 and Physics 2. While all students shift towards less favorable attitudes and beliefs, female students shift more negatively than male students. In Physics 1, the largest differences in shifts relate to students' attitudes and beliefs about problem-solving, while in Physics 2, females have significantly larger shifts compared to males in their personal interest in physics. The introductory physics courses are negatively impacting all students' attitudes and beliefs about physics, but they are more negatively impacting female students.

We also examined retention from Physics 1 to Physics 2, to see what happens to students after they complete Physics 1. As reported in Chapter 3, while there are no gender differences in the rates of all students continuing or not continuing from Physics 1 to Physics 2, there are gender differences in the rates for physics majors. Female physics majors are more likely to drop out of the physics major after Physics 1 (by either changing majors or not taking Physics 2) than males. In Chapters 2 and 3, we characterized gender differences in student performance (both prior to and in the introductory class), student attitudes and beliefs, and retention in the introductory sequence.

2. *Can gender differences in student performance be accounted for by factors other than gender, namely, background differences of males and females?*

In Chapters 2 and 3, we used regression modeling to control for a variety of factors that differ by gender, including prior physics and mathematics understanding, high school courses taken and performance, and incoming attitudes and beliefs, as well as demographic factors. Controlling for student background allows us to theoretically compare the performance of males and females who are equally prepared walking into the introductory physics course. Between 60% and 70% of the gender difference in post-course conceptual survey performance can be accounted for by differences in males and females' prior physics and mathematics performance (as measured by conceptual survey pre-test and standardized math test scores) and their incoming attitudes and beliefs about physics. That is, the majority of the gender gap that we observe at the end of Physics 1 or Physics 2 can be attributed to the pre-course underpreparation of female students compared to male students.

Several variables that we hypothesized would be important predictors of student performance, did not make it into our regression models. Background factors, such as high school GPA and high school physics courses taken, as well as students' declared major and ethnicity were not significant predictors of students' post-course conceptual survey performance once the variables in our final models were already included. That is not to say that these variables do not impact student performance, but rather, they do not provide any additional predictive power beyond prior physics and mathematics understanding and incoming attitudes and beliefs about physics. It is likely that these variables are highly correlated with student background, for instance, physics and engineering majors have higher pre-course physics scores than nonscience majors, which is why they were not significant predictors in our models.

3. *What role do underlying psychological factors, such as students' physics self-confidence (self-efficacy) and sense of being a physics person (physics identity), play in student performance and in the gender differences in student performance?*

Having identified cognitive factors, like prior physics and mathematics understanding, that impact student performance and the gender gap, in Chapters 4 and 5 we investigated the psychological factors that underpin student performance and learning in the course. These psychological factors, specifically self-efficacy and identity, provide potential mechanisms for how gender disparities arise in the introductory physics courses. Using Bandura's self-efficacy theory [4], we explored whether males and females came into the introductory courses with different levels of physics self-efficacy, how their self-efficacy developed differently during the introductory courses, and whether self-efficacy was a useful predictor of student performance in the course. Females report being less confident than males about completing the tasks necessary for success coming into the physics courses, though the differences are largely related to tasks that require performance, such as taking tests or discussing physics with peers, as opposed to using mathematics in physics, working with other students, or getting help in the course. Regression models showed that students' pre-course *performance* self-efficacy was a significant predictor of their exam and conceptual survey performance, even when already accounting for student background. This indicates that both students' prior performance and their confidence about succeeding (likely based on their prior performance) impact their performance in the introductory physics courses. Differences in males and females' pre-course *performance* self-efficacy accounted for an additional 12% of the gender gap in post-course conceptual survey scores, once student background had already accounted for about 70% of the gender gap.

The development of students' self-efficacy in the introductory physics courses was examined through the four sources of self-efficacy. Students were asked to report their

perceptions of their mastery experiences, vicarious experiences, verbal persuasion, and emotional responses. Females reported more negative interpretations of their experiences than males for all four of the sources of self-efficacy. As predicted by self-efficacy theory, all four of these sources were significant predictors of students' post-course self-efficacy. Given females' more negative interpretations of their experiences and the correlations between the sources and self-efficacy, it is not surprising that female students had lower post-course self-efficacy than males, not just for tasks related to performance, but also in using mathematics in physics and in getting help in the physics course. In this study of self-efficacy in physics we found that males and females interpret their experiences in the introductory physics courses differently, those interpretations result in differences in physics self-efficacy beliefs, which then contributes to the differences in performance between males and females in the introductory physics courses.

In Chapter 5 we turned to a second psychological factor that underpins student performance and student learning in the introductory physics courses. Using a model of physics identity developed in prior literature [5,6], we investigated students' recognition of themselves and their perceptions of others' recognition of them as a physics person. Male students report having a greater sense of physics identity than female students. Male students are more likely to report that they see themselves and are seen by others (peers, teachers, and parents) as a physics person. When asked to report why they do or do not see themselves as a physics person, most students gave reasons related to their interest in physics or their ability to do physics. Both male and female students in Physics 1 report that their self-identity decreases over the course of the semester, while students' in Physics 2 report no change in their perception of themselves as a physics person. While students' self-identity was a significant predictor of students' performance on its own, it was not as strong a predictor as other variables, such as background physics and

mathematics understanding and pre-course self-efficacy, and was not significant when these other variables were included in the model. Though identity does not appear to impact student performance in the introductory physics courses, we suspect, and prior literature suggests, that it is an important factor in student retention in physics and needs to be further studied.

4. *Does identity threat (the fear of being devalued based on a group identity) negatively impact females' performance in the introductory physics course? Can the performance differences between males and females be reduced or even eliminated with a psychological, self-affirmation intervention?*

In our final study (Chapter 6), we implemented a self-affirmation intervention in Physics 1 to enhance the exam and conceptual survey performance of female students. The effectiveness of the brief writing intervention was tested using a randomized experimental design. The self-affirmation intervention was successful in reducing the gender gap both in students' in-class exam scores and their conceptual post-test scores. The reduction in the gender gap was primarily due to affirmed females' increased performance compared to females in the control group. Further, the effect of the intervention was moderated by students' endorsement of the stereotype that men are expected to do better in physics than women. Stereotype endorsement and performance on exams and the conceptual survey were negatively related for female students, unless they had completed the self-affirmation exercise, in which case they were buffered from the negative impact of stereotype endorsement. This initial study suggests that identity threat may in fact be impacting student performance in the introductory physics courses, and that self-affirmation may be a promising way to reduce or even eliminate gender differences in student performance.

Following up on the initial study, we conducted a replication study of the self-affirmation intervention in a later semester when a different instructor was teaching the course. Though we cannot make any statistical claims regarding the efficacy of the intervention in this semester due to the limited sample sizes of students, we did not observe the same trends as in the initial study. That is, the gender gap among affirmed students was not reduced compared to the gender gap among control students. The replication study makes clear that the intervention is not a “silver bullet,” and likely needs to be implemented in conjunction with classroom practices that promote student learning and in a classroom environment that is supportive of all students, as has been suggested by prior researchers [7].

Taken together, this thesis works towards building a fine-grained model of student performance in the introductory physics courses. So far, this model includes background cognitive factors, such as prior physics and mathematics understanding, prior psychological factors, such as students’ attitudes and beliefs about physics and students’ physics self-efficacy, and social factors, such as gender and possibly identity threat, as suggested by the results of the self-affirmation intervention. Additionally, we have worked towards developing a model of gender differences in performance in the introductory physics courses. One goal of this work was to understand gender differences in performance in physics. The findings of this thesis suggest that there are at least three contributing factors: 1) differences in the background of males and females, i.e., physics and mathematics understanding and attitudes and beliefs about physics, which is the largest contributor, 2) differences in the pre-course self-efficacy of males and females, and 3) identity threat. Similar to Valian’s notion of “accumulated (dis)advantage” [8] or Tatum’s “smog of bias” [9], we find that it is not one factor that is responsible for the gender disparities in physics participation and performance. Rather, female students consistently fall

behind male students across a variety of factors as they move through the introductory physics sequence. These small, consistent differences can build up and accrue over time to result in large disparities. Establishing this model of gender differences in performance was a first step towards increasing females' participation and performance in physics, and can be used to guide future interventions to address the disparities.

II. Implications

Having summarized the major findings of this thesis, we now discuss the implications of the results for the structure and pedagogy of the introductory physics class. The majority of the gender gap in post-course performance can be attributed to pre-course preparation differences of males and females. If we want males and females to leave the introductory courses with equal levels of physics understanding, this preparation gap must be addressed. One view is that it needs to be addressed in the high schools. Since there are differences when students walk in the college physics classroom, we need to go back to where those differences developed and focus effort there. Another perspective, one that we take, is that it is our responsibility to teach the students we have. We need to find ways to teach to the under-prepared students in the course (both male and female), especially as this is an introductory course with no prior physics instruction required. One option is to create a co-seminar that students would take in conjunction with the introductory physics course. This co-seminar could focus on improving the mathematical skills of under-prepared students, or simply allow more time for students to interact with peers and instructors. Another option would be to create a separate Physics 1 course that covers the same material but in twice the time (i.e., twice the lecture and recitation time per week), giving underprepared students time to catch up on the background that they are missing and allowing more time to learn the current course material. This model has been successful at improving the

performance of mathematically underprepared students at other universities [10] and in the Applied Mathematics department at CU [11]. Alternatively, the curriculum of the introductory courses could be changed, such that it is aimed at students with less background in physics. Researchers have found that courses that cover less material but go into greater depth help students of both genders learn more [12]. While there are many possible ways to address the preparation gap of males and females, all of these approaches need to be investigated.

Another aspect of the introductory courses that should be examined is assessment. Currently, in-class exams that are high-stakes, timed, individual, and mostly multiple-choice make up 60% to 75% of students' course grades. Findings from this thesis suggest that there are factors other than student understanding of physics that contribute to students' scores on these exams. Psychological factors, such as self-efficacy and identity threat, disproportionately disadvantage female students on such tasks. Female students report feeling more anxiety and stress about exams than males, which negatively impacts their exam performance. Additionally, female students do better on homework (one indication of students' understanding of physics), but homework is less valued in the course than exams. Changing the structure of course grades to make exams less heavily weighted could alleviate anxiety and stress for all students, allowing them to perform better. In addition, incorporating psychological interventions that have been shown to boost exam performance, like self-affirmation [7] or others [13], could be beneficial. In light of our findings, the use of high-stakes tests as the primary assessment tool in the introductory physics courses needs to be reconsidered, as it disproportionately hurts female students.

Given our findings that females are disproportionately uninterested in physics both before they enter the introductory physics courses and after they complete them, perhaps the curriculum

needs to be adapted to the interests of female students. Research shows that the science topics that female students are interested in are often also of interest to male students, though the reverse is not always true [14]. Incorporating more topics of broad interest in the introductory courses could also help to show students how physics connects to their specific, non-physics interests or their lives outside of class. Prior literature on increasing girls' interest in science suggests that making connections to real life is important [15,16,17]. Additionally, research shows that discussing current relevant science and discussing the benefits of being a scientist can positively impact students' science identities [6,16]. Restructuring and rewriting the curriculum of the introductory physics courses would be a huge undertaking. Before taking on the challenge (and probably regardless of whether the curriculum is changed or not), it is important to discuss and come to consensus about what the purpose of the introductory courses is, who we are teaching these courses to, and what we want students to get out of them. If we, as educators and researchers, are committed to increasing the performance and participation of females in physics, then these discussions need to happen with an explicit focus on the experiences, interests, and needs of female students.

III. Directions for Future Research

In this thesis, we identified that the largest contributor to gender differences in performance in the introductory physics courses is differences in the background and preparation of male and female students. One direction for future research is to investigate where these background differences come from. Possible research questions include:

- Are males and females taking the same kinds of high school physics courses? Do males and females who take the same kind of high school class come into the introductory college courses with the same background physics understanding?

- Do gender gaps in conceptual physics understanding exist in high school physics courses, both at the beginning and at the end?
- What experiences did students have in their high school physics (or science) courses that led to their attitudes and beliefs, self-efficacy, and identity coming into the introductory college courses?

Throughout this thesis, we identified individual semesters for which the overall trends did not hold. Future research could focus on faculty and the choices that they make throughout their instruction, to identify practices that reduce or enhance gender differences in the introductory physics courses. Possible research questions are:

- How can the classroom practices of faculty be characterized and distinguished? How can classroom norms and the classroom culture established by the instructor be characterized?
- Do gender differences in student performance correlate with instructors' practices and choices during instruction, or with specific classroom norms?
- Can particular instructor practices be identified that improve or harm female performance or attitudes in the introductory courses?

In addition to looking back at where gender differences came from and investigating the introductory courses further, it is also important to look forward as students continue through the physics major.

- Do gender differences exist in the upper-division physics courses?
- What experiences do female students have as they work through the physics major that contribute to their persistence in physics or their decision to leave physics?

- What are the characteristics of successful female physics majors? What did they look like in the introductory physics courses? Can we identify the introductory female students who are likely to become physics majors?

In looking further along the trajectory as students become physics majors, we suspect that identity and students' identity development will be critical factors. Potential research questions include:

- How are the physics identities that male and female students develop different? Does the physics major curriculum and pedagogy value particular physics identities?
- What kinds of experiences are important and productive for developing students' physics identities? How can we provide more of those experiences?
- Does the physics identity that students develop align with the identity of practicing physicists?

Lastly, in this thesis, we tested the effectiveness of a self-affirmation intervention. Further research is needed (and is currently underway by our psychology colleagues) to understand:

- What are the conditions under which self-affirmation is beneficial?
- By what mechanism does the self-affirmation intervention work? Does the intervention alleviate stress to allow more space in working memory? Does it change how students interact and engage in the course? Does it impact students' identity or self-efficacy?
- Is the self-affirmation beneficial for other student populations and in other science courses where there are documented gender differences?

In addition to the self-affirmation intervention, there is other work to be done on interventions:

- What other interventions have been developed to target female participation and performance in science? How do these interventions align with our statistical model of student performance and gender differences in the course? What is the impact of these interventions?
- What interventions have been developed to target background differences of males and females? Are they effective at overcoming the pre-course gender differences that are observed in the introductory physics courses?
- What are the strengths of female students in the introductory physics courses? Can interventions be developed that draw upon the particular strengths of female students?

A lot of work is needed before we understand why fewer females participate in physics. While we have focused on introductory college physics in this thesis, there is work to be done (and is being done) at all levels along the academic and career trajectories. Underlying all of this work is the culture of physics, the practices, attitudes, identities, beliefs, values, and goals of the community of physicists. We suspect that critical analysis of the culture of physics, how it supports and disadvantages females, and how it can be changed to be more inclusive of females, is necessary in order to achieve our goal of equal participation of males and females in physics. As physics educators and researchers, we all must work towards creating classrooms, laboratories, communities, and cultures that are inclusive and supportive of all students, regardless of gender (as well as race, ethnicity, religion, sexual orientation, etc.), and allow all students to participate equally.

References (Chapter 7)

- [1] E. Mazur, *Peer Instruction: A Users Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- [2] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- [3] CAPA, <http://www.lon-capa.org/>; Mastering Physics, <http://www.masteringphysics.com/>.
- [4] A. Bandura, *Self-Efficacy: The Exercise of Control* (Freeman, New York, 1997).
- [5] H. B. Carlone and A. Johnson, Understanding the science experiences of successful women of color: Science identity as an analytic lens, *Journal of Research in Science Teaching* **44**(8), 1187 (2007).
- [6] Z. Hazari, G. Sonnert, P. M. Sadler and M. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, *Journal of Research in Science Teaching* **47**(8), 978 (2010).
- [7] G. L. Cohen, J. Garcia, N. Apfel and A. Master, Reducing the racial achievement gap: A social-psychological intervention, *Science* **313**, 1307 (2006).
- [8] V. Valian, *Why So Slow: The Advancement of Women* (The MIT Press, Cambridge, 1999).
- [9] B. D. Tatum, *Why are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race* (Basic Books, New York, 1997).
- [10] S. W. Brahmia, “Improving learning for underrepresented groups in physics for engineering majors” in *Proceedings of the 2008 Physics Education Research Conference*, edited by C. Henderson, M. Sabella, and L. Hsu (AIP Conference Proceedings, New York, 2008) p. 7.
- [11] M. F. Nelson, A reform approach to calculus instruction: Effects on retention, grades and understanding, Ph.D. dissertation, University of Colorado at Boulder (2005).
- [12] R. H. Tai, and P. M. Sadler, Gender differences in introductory undergraduate physics performance: university physics versus college physics in the USA, *International Journal of Science Education* **23**, 1017 (2001).
- [13] G. Ramirez and S. L. Beilock, Writing about testing worries boosts exam performance in the classroom, *Science* **331**, 211 (2011).
- [14] P. Häussler and L. Hoffman, An intervention study to enhance girls’ interest, self-concept, and achievement in physics classes, *Journal of Research in Science Teaching* **39**(9), 870 (2002).

- [15] D. Halpern, J. Aronson, N. Reimer, S. Simpkins, J. Star and K. Wentzel, *Encouraging Girls in Math and Science* (NCER 2007-2003). Washington DC: National Center for Education Research, Institute of Education Sciences, U. S. Department of Education (2007). Retrieved from <http://ncer.ed.gov>.
- [16] *Teaching the Majority: Breaking the Gender Barrier in Science, Mathematics, and Engineering*, edited by S. V. Rosser (Teacher's College Press, New York, 1995).
- [17] C. S. Hulleman and J. M. Harackiewicz, Promoting interest and performance in high school science classes, *Science* **326**, 1410 (2009).

BIBLIOGRAPHY

- W. K. Adams, K. K. Perkins, N. Podolefsky, M. Dubson, N. D. Finkelstein and C. E. Wieman, A new instrument for measuring student beliefs about physics and learning physics: the Colorado Learning Attitudes about Science Survey, *Physical Review Special Topics - Physics Education Research* **2**, 010101 (2006).
- M. Adamuti-Trache and L. Andres, Embarking on and persisting in scientific fields of study: Cultural capital, gender, and curriculum along the science pipeline, *International Journal of Science Education* **30**(12), 1557 (2008).
- L. S. Aiken and S. G. West, *Multiple Regression: Testing and Interpreting Interactions* (Sage, Thousand Oaks, CA, 1991).
- N. Ambady, M. Shih, A. Kim and T. L. Pittinsky, Stereotype susceptibility in children: Effects of identity activation on quantitative performance, *Psychological Science* **12**, 385 (2001).
- America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science (COMPETES) Act, Public Law 110-69, 121 Stat. 572 (2007).
- America COMPETES Reauthorization Act of 2010, Public Law 111-358, 124 Stat. 3982 (2011).
- American Physical Society, Committee on the Status of Women in Physics, *Gender Equity Report* (unpublished, 2007). Available from <http://www.aps.org/programs/women/workshops/gender-equity/>.
- APS Panel on Public Affairs, Committee on Energy and Environment, *Readiness of the U.S. Nuclear Workforce for 21st Century Challenges* (2008). Available at <http://www.aps.org/policy/reports/popa-reports/index.cfm>.
- APS/AAPT Doubling Initiative Mission Statement (2007). Available at <http://www.aps.org/programs/education/undergrad/faculty/double-mission.cfm>.
- J. Aronson, C. B. Fried and C. Good, Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence, *Journal of Experimental Social Psychology* **38**, 113 (2002).
- J. Aronson and M. Inzlicht, The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African American college students, *Psychological Science* **15**, 829 (2004).

- J. Aronson, M. Lustina, C. Good, K. Keough, C. M. Steele and J. Brown, When white men can't do math: Necessary and sufficient factors in stereotype threat, *Journal of Experimental Social Psychology* **35**, 29 (1999).
- A. Bandura, *Self-Efficacy: The Exercise of Control* (Freeman, New York, 1997).
- A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, *Psychological Review* **84**(2), 191 (1977).
- A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice Hall, Englewood Cliffs, NJ, 1986).
- A. C. Barton, Teaching science with homeless children: Pedagogy, representation, and identity, *Journal of Research in Science Teaching* **35**(4), 379 (1998).
- A. C. Barton, J. L. Ermer, T. A. Burkett and M. D. Osborne, *Teaching Science for Social Justice* (Teachers College Press, New York, 2003).
- A. C. Barton and K. Yang, The culture of power and science education: Learning from Miguel, *Journal of Research in Science Teaching* **37**(8), 871 (2000).
- S. L. Beilock, R. J. Rydell, and A. R. McConnell, Stereotype threat and working memory: Mechanisms, alleviation and spillover, *Journal of Experimental Psychology: General* **136**(2), 256 (2007).
- M. F. Belenky, B. M. Clinchy, N. R. Goldberger and J. M. Tarule, *Women's Ways of Knowing: The Development of Self, Voice, and Mind* (Basic Books, New York, 1986).
- R. A. Berk, *Regression Analysis: A Constructive Criticism* (SAGE Publications, Thousand Oaks, CA, 2004), p. 212.
- N. E. Betz and G. Hackett, The relationship of career-related self-efficacy expectations to perceived career options in college women and men, *Journal of Counseling Psychology* **28**(5), 399 (1981).
- N. E. Betz and G. Hackett, The relationship of mathematics self-efficacy expectations to the selection of science-based college majors, *Journal of Vocational Behavior* **23**, 329 (1983).
- J. C. Blickenstaff, Women and science careers: leaky pipeline or gender filter? *Gender and Education* **17**(4), 369 (2005).
- J. Boaler, The development of disciplinary relationships: knowledge, practice, and identity in mathematics classrooms, *For the Learning of Mathematics* **22**(1), 42 (2002).

- J. Boaler and J. Greeno, "Identity, agency, and knowing in mathematics worlds" in *Multiple Perspectives on Mathematics Teaching and Learning* edited by J. Boaler (Ablex Publishing, Westport, CT, 2000) p. 171.
- S. W. Brahmia, "Improving learning for underrepresented groups in physics for engineering majors" in *Proceedings of the 2008 Physics Education Research Conference*, edited by C. Henderson, M. Sabella, and L. Hsu (AIP Conference Proceedings, New York, 2008) p. 7.
- S. Brahmia and E. Etkina, Switching students on to science, *Journal of College Science Teaching* **31**(3), 183 (2001).
- E. Brewster, Modeling theory applied: Modeling Instruction in introductory physics, *American Journal of Physics* **76**(12), 1155 (2008).
- E. Brewster, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá, Toward equity through participation in Modeling Instruction in introductory university physics, *Physical Review Special Topics – Physics Education Research* **6**, 010106 (2010).
- N. Brickhouse, Bringing in the outsiders: Reshaping the science of the future, *Journal of Curriculum Studies* **26**(4), 401 (1994).
- N. W. Brickhouse, Embodying science: A feminist perspective on learning, *Journal of Research in Science Teaching* **38**(3), 282 (2001).
- N. W. Brickhouse, P. Lowery and K. Schultz, What kind of a girl does science? The construction of school science identities, *Journal of Research in Science Teaching* **37**(5), 441 (2000).
- N. W. Brickhouse and J. T. Potter, Young women's scientific identity formation in an urban context, *Journal of Research in Science Teaching* **38**(8), 965 (2001).
- S. L. Britner and F. Pajares, Sources of science self-efficacy beliefs of middle school students, *Journal of Research in Science Teaching* **43**(5), 485 (2006).
- E. Brogt, D. Sabers, E. E. Prather, G. L. Deming, B. Hufnagel and T. F. Slater, Analysis of the astronomy diagnostic test, *Astronomy Education Review* **6**, 25 (2007).
- J. S. Brotman and F. M. Moore, Girls and science: A review of four themes in the science education literature, *Journal of Research in Science Teaching* **45**(9), 971 (2008).
- J. L. Brown and R. A. Josephs, A burden of proof: Stereotype relevance and gender differences in math performance, *Journal of Personality and Social Psychology* **76**, 246 (1999).
- A. Bug, Has feminism changed physics? *Journal of Women in Culture and Society* **28**, 881 (2003).

CAPA, <http://www.lon-capa.org/>.

H. B. Carlone, Innovative science within and against a culture of “achievement”, *Science Education* **87**(3), 307 (2003).

H. B. Carlone, The cultural reproduction of science in reform-based physics: Girls’ access, participation, and resistance, *Journal of Research in Science Teaching* **41**(4), 392 (2004).

H. B. Carlone and A. Johnson, Understanding the science experiences of successful women of color: Science identity as an analytic lens, *Journal of Research in Science Teaching* **44**(8), 1187 (2007).

S. E. Carrell, M. E. Page and J. E. West, Sex and science: How professor gender perpetuates the gender gap, NBER Working Paper 14959 (Cambridge, MA; May 2009). Available from <http://www.nber.org/papers/w14959>.

G. Carter, S. L. Westbrook and C. D. Thompkins, Examining science tools as mediators of students’ learning about circuits, *Journal of Research in Science Teaching* **36**(1), 89 (1999).

A. M. L. Cavallo, M. Rozman, and W. H. Potter, Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors, *School Science and Mathematics* **104**, 288 (2004).

S. J. Ceci and W. M. Williams, Sex differences in math-intensive fields, *Current Directions in Psychological Science* **19**(5), 275 (2010).

S. J. Ceci, W. M. Williams and S. M. Barnett, Women’s underrepresentation in science: Sociocultural and biological considerations, *Psychological Bulletin* **135**(2), 218 (2009).

D. Cervone and P. K. Peake, Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior, *Journal of Personality and Social Psychology* **50**(3), 492 (1986).

R. Chabay and B. Sherwood, Restructuring the introductory electricity and magnetism course, *American Journal of Physics* **74**, 329 (2006).

M. M. Chemers, L. Hu and B. F. Garcia, Academic self-efficacy and first-year college student performance and adjustment, *Journal of Educational Psychology* **93**(1), 55 (2001).

P. W. U. Chinn, Asian and Pacific Islander women scientists and engineers: A narrative explanation of model minority, gender, and racial stereotypes, *Journal of Research in Science Teaching* **39**(4), 302 (2002).

- A. Cleaves, The formation of science choices in secondary school, *International Journal of Science Education* **27**(4), 471 (2005).
- P. Cobb, Mathematics, literacies, and identity, *Reading Research Quarterly* **39**, 332 (2004).
- G. L. Cohen, J. Aronson and C. M. Steele, When beliefs yield to evidence: Reducing biased evaluation by affirming the self, *Personality and Social Psychology Bulletin* **26**, 1151 (2000).
- G. L. Cohen and J. Garcia, "I am us": Negative stereotypes as collective threats, *Journal of Personality and Social Psychology* **89**, 566 (2005).
- G. L. Cohen and J. Garcia, Identity, belonging, and achievement: A model, interventions, and implications, *Current Directions in Psychological Science* **17**(6), 365 (2008).
- G. L. Cohen, J. Garcia, N. Apfel and A. Master, Reducing the racial achievement gap: A social-psychological intervention, *Science* **313**, 1307 (2006).
- G. L. Cohen, J. Garcia, V. Purdie-Vaughns, N. Apfel and P. Brzustoski, Recursive processes in self-affirmation: Intervening to close the minority achievement gap, *Science* **324**, 400 (2009).
- B. Cole, K. Matheson and H. Anisman, The moderating role of ethnic identity and social support on relations between well-being and academic performance, *Journal of Applied Social Psychology* **37**, 592 (2007).
- College Board, *Advanced Placement Summary Information* (The College Board, New York, 2005).
- W. J. Conover, *Practical Nonparametric Statistics* (John Wiley & Sons, New York, 1999).
- E. D. Cordero, S. H. Porter, T. Israel and M. T. Brown, Math and science pursuits: A self-efficacy intervention comparison study, *Journal of Career Assessment* **18**(4), 362 (2010).
- M. Dancy, "The myth of gender neutrality" in *Proceedings of the 2003 Physics Education Research Conference*, edited by J. Marx, S. Franklin, and K. Cummings (AIP Conference Proceedings, New York, 2004) p. 31.
- A. T. Danielsson and C. Linder, Learning in physics by doing laboratory work: Towards a new conceptual framework, *Gender and Education* **21**(2), 129 (2009).
- L. Ding, R. Chabay, B. Sherwood and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Physical Review Special Topics – Physics Education Research* **2**, 010105 (2006).

- L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, *Physical Review Special Topics – Physics Education Research* **5**, 020103 (2009).
- J. Docktor and K. Heller, “Gender differences in both Force Concept Inventory and introductory physics performance” in *Proceedings of the 2008 Physics Education Research Conference*, edited by C. Henderson, M. Sabella and L. Hsu (AIP Conference Proceedings, New York, 2008) p. 15.
- Domestic Policy Council, Office of Science and Technology Policy, *American Competitiveness Initiative: Leading the World in Innovation* (2006). Available at www.nsf.gov/attachments/108276/public/ACI.pdf
- M. A. Eisenhart and E. Finkel, *Women’s Science: Learning and Succeeding from the Margins* (The University of Chicago Press, Chicago, 1998).
- H. Fencl and K. Scheel, Engaging students: An examination of the effects of teaching strategies on self-efficacy and course climate in a nonmajors physics course, *Journal of College Science Teaching* **35**(1), 20 (2005).
- N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: implementing tutorials in introductory physics, *Physical Review Special Topics - Physics Education Research* **1**, 010101 (2005).
- J. P. Gee, Identity as an analytic lens for research in education, *Review of Research in Education* **25**, 99 (2000).
- E. Gillibrand, P. Robinson, R. Brawn and A. Osborn, Girls’ participation in physics in single sex classes in mixed schools in relation to confidence and achievement, *International Journal of Science Education* **21**(4), 349 (1999).
- C. Good, J. Aronson and J. A. Harder, Problems in the pipeline: Stereotype threat and women’s achievement in high-level math courses, *Journal of Applied Developmental Psychology* **29**, 17 (2008).
- C. Good, J. Aronson and M. Inzlicht, Improving adolescents’ standardized test performance: An intervention to reduce the effects of stereotype threat, *Applied Developmental Psychology* **24**, 645 (2003).
- L. Guiso, F. Monte, P. Sapienza and L. Zingales, Culture, gender, and math, *Science* **320**, 1164 (2008).
- A. Gungor, A. Eryilmaz and T. Fakioglu, The relationship of freshman’s physics achievement and their related affective characteristics, *Journal of Research in Science Teaching* **44**(8), 1036 (2007).

- G. Hackett, The role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis, *Journal of Counseling Psychology* **32**(1), 47 (1985).
- G. Hackett and N. E. Betz, A self-efficacy approach to the career development of women, *Journal of Vocational Behavior* **18**(3), 326 (1981).
- G. Hackett and N. E. Betz, An exploration of the mathematics self-efficacy/mathematics performance correspondence, *Journal for Research in Mathematics Education* **20**(3), 261 (1989).
- R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *American Journal of Physics* **66**, 64 (1998).
- R. R. Hake, "Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pre-test scores on mathematics and spatial visualization" in *Proceedings of the 2002 Physics Education Research Conference*, edited by S. Franklin, K. Cummings and J. Marx (AIP Conference Proceedings, New York, 2003).
- D. Halpern, J. Aronson, N. Reimer, S. Simpkins, J. Star and K. Wentzel, *Encouraging Girls in Math and Science* (NCER 2007-2003). Washington DC: National Center for Education Research, Institute of Education Sciences, U. S. Department of Education (2007). Retrieved from <http://ncer.ed.gov>
- N. Z. Hampton and E. Mason, Learning disabilities, gender, sources of self-efficacy, self-efficacy beliefs, and academic achievement in high school students, *Journal of School Psychology* **41**, 101 (2003).
- S. Harding, *Whose Science? Whose Knowledge? Thinking from Women's Lives* (Cornell University Press, Ithaca, NY, 1991).
- P. Häussler and L. Hoffman, An intervention study to enhance girls' interest, self-concept, and achievement in physics classes, *Journal of Research in Science Teaching* **39**(9), 870 (2002).
- Z. Hazari, P. M. Sadler and R. H. Tai, Gender differences in the high school and affective experiences of introductory college physics students, *The Physics Teacher* **46**, 423 (2008).
- Z. Hazari, G. Sonnert, P. M. Sadler and M. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, *Journal of Research in Science Teaching* **47**(8), 978 (2010).

- Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: the influence of high school physics preparation and affective factors, *Science Education* **91**, 847 (2007).
- E. Hazel, P. Logan and P. Gallagher, Equitable assessment of students in physics: importance of gender and language background, *International Journal of Science Education* **19**, 381 (1997).
- A. F. Heckler, Measuring student learning by pre and post testing: absolute gain vs. normalized gain, Unpublished. Can be found at: <http://www.mps.ohio-state.edu/Personnel/Heckler/TauVsGpaper.final.pdf>
- P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *American Journal of Physics* **60**, 637 (1992).
- P. Heller, R. Keith and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *American Journal of Physics* **60**, 627 (1992).
- M. K. Hemenway, W. J. Straits, R. R. Wilke and B. Hufnagel, Educational research in an introductory astronomy course, *Innovative Higher Education* **26**(4), 271 (2002).
- D. Hestenes, Toward a modeling theory of instruction, *American Journal of Physics* **55**, 440 (1987).
- D. Hestenes, M. Wells and G. Swackhamer, Force concept inventory, *The Physics Teacher* **30**, 141 (1992).
- D. C. Holland and M. A. Eisenhart, *Educated in Romance: Women, Achievement, and College Culture* (University of Chicago Press, Chicago, IL, 1990).
- B. Hufnagel, G. L. Deming, J. M. Landato and A. K. Hodari, The effect of stereotype threat on undergraduates in an introductory astronomy class, *Journal of Women and Minorities in Science and Engineering* **10**, 89 (2004).
- G. Hughes, Exploring the availability of student scientist identities within curriculum discourse: An anti-essentialist approach to gender-inclusive science, *Gender and Education* **13**(3), 275 (2001).
- C. S. Hulleman and J. M. Harackiewicz, Promoting interest and performance in high school science classes, *Science* **326**, 1410 (2009).
- M. A. Hutchison, D. K. Follman, M. Sumpter and G. M. Bodner, Factors influencing the self-efficacy beliefs of first-year engineering students, *Journal of Engineering Education* **95**(1), 39 (2006).

- J. S. Hyde, S. M. Lindberg, M. C. Linn, A. B. Ellis and C. C. Williams, Gender similarities characterize math performance, *Science* **321**, 494 (2008).
- M. Inzlicht and T. Ben-Zeev, A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males, *Psychological Science* **11**, 365 (2000).
- R. Ivie and K. N. Ray, *Women in Physics and Astronomy, 2005*, AIP Report No. R-430.02 (College Park, MD, 2005). Available from www.aip.org/statistics.
- A. C. Johnson, Unintended consequences: How science professors discourage women of color, *Science Education* **91**(5), 805 (2007).
- M. G. Jones, L. Brader-Araje, L. Carboni, G. Carter, M. Rua, E. Banilower, et al., Tool time: Gender and students' use of tools, controls, and authority, *Journal of Research in Science Teaching* **37**, 760 (2000).
- J. Jovanovic and S. King, Boys and girls in the performance-based science classroom: Who's doing the performing? *American Educational Research Journal* **35**, 477 (1998).
- C. M. Judd and G. H. McClelland, *Data Analysis: A Model Comparison Approach* (Harcourt Brace Jovanovich, New York, 1989).
- S. K. Kachigan, *Multivariate Statistical Analysis: A Conceptual Introduction* (Radius Press, New York, 1991).
- J. B. Kahle, Will girls be left behind? Gender differences and accountability, *Journal of Research in Science Teaching* **41**(10), 961 (2004).
- J. B. Kahle, A. Anderson and A. Damnjanovic, A comparison of elementary teacher attitudes and skills in teaching science in Australia and the United States, *Research in Science Education* **21**, 208 (1991).
- J. B. Kahle and J. Meece, "Research on gender issues in the classroom" in *Handbook of Research on Science Teaching and Learning*, edited by D. Gabel (National Science Teachers Association, Washington, DC, 1994) p. 542.
- J. B. Kahle, L. H. Parker, L. J. Rennie and D. Riley, Gender differences in science education: Building a model, *Educational Psychology* **28**(4), 379 (1993).
- A. Kelly, The construction of masculine science, *British Journal of Sociology of Education* **6**, 133 (1985).
- J. Kenway and A. Gough, Gender and science education in schools: A review "with attitude," *Studies in Science Education* **31**(1), 1 (1998).

- M. A. Kohlmyer, M. D. Caballero, R. Catrambone, R. W. Chabay, L. Ding, M. P. Haugan, M. J. Marr, B. A. Sherwood and M. F. Schatz, Tale of two curricula: The performance of 2000 students in introductory electromagnetism, *Physical Review Special Topics – Physics Education Research* **5**, 020105 (2009).
- L. J. Kray, A. D. Galinsky and L. Thompson, Reversing the gender gap in negotiations: An exploration of stereotype regeneration, *Organizational Behavior and Human Decision Processes* **87**, 386 (2002).
- J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation* (Cambridge University Press, Cambridge, 1991).
- P. Laws, P. Rosborough and R. Poodry, Women's responses to an activity-based introductory physics program, *American Journal of Physics* **67**, S32, (1999).
- J. D. Lee, More than ability: Gender and personal relationships influence science and technology involvement, *Sociology of Education* **75**, 349 (2002).
- J. D. Lee, Which kids can “become” scientists? Effects of gender, self-concepts, and perceptions of scientists, *Social Psychology Quarterly* **61**(3), 199 (1998).
- R. W. Lent and S. D. Brown, On conceptualizing and assessing social cognitive constructs in career research: A measurement guide, *Journal of Career Assessment* **14**, 12 (2006).
- R. W. Lent, S. D. Brown and G. Hackett, Toward a unifying social cognitive theory of career and academic interest, choice, and performance, *Journal of Vocational Behavior* **45**, 79 (1994).
- R. W. Lent, S. D. Brown and K. C. Larkin, Comparison of three theoretically derived variables in predicting career and academic behavior: Self-efficacy, interest congruence, and consequence thinking, *Journal of Counseling Psychology* **34**(3), 293 (1987).
- R. W. Lent, S. D. Brown and K. C. Larkin, Relation of self-efficacy expectations to academic achievement and persistence, *Journal of Counseling Psychology* **31**(3), 356 (1984).
- R. W. Lent, S. D. Brown and K. C. Larkin, Self-efficacy in the prediction of academic performance and perceived career options, *Journal of Counseling Psychology* **33**(3), 265 (1986).
- R. W. Lent, F. G. Lopez and K. J. Bieschke, Mathematics self-efficacy: Sources and relation to science-based career choice, *Journal of Counseling Psychology* **38**(4), 424 (1991).
- R. W. Lent, H. Sheu, D. Singley, J. A. Schmidt, L. C. Schmidt and C. S. Gloster, Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students, *Journal of Vocational Behavior* **73**, 328 (2008).

- B. Levy, Improving memory in old age through implicit self stereotyping, *Journal of Personality and Social Psychology* **71**, 1092 (1996).
- M. C. Linn and J. S. Hyde, Gender, mathematics, and science, *Educational Researcher* **18**(8), 17 (1989).
- M. C. Linn and A. C. Peterson, "Facts and assumptions about the nature of sex differences" in *Handbook for Achieving Sex Equity Through Education*, edited by S. S. Klein (Johns Hopkins University Press, Baltimore, MD, 1985) p. 53.
- R. G. Lomax, *An Introduction to Statistical Concepts for Education and Behavioral Sciences* (Lawrence Erlbaum Associates, Mahwah, NJ, 2001).
- F. G. Lopez, R. W. Lent, S. D. Brown and P. A. Gore Jr., Role of social-cognitive expectations in high school students' mathematics-related interest and performance, *Journal of Counseling Psychology* **44**(1), 44 (1997).
- M. Lorenzo, C. Crouch and E. Mazur, Reducing the gender gap in the physics classroom, *American Journal of Physics* **74**, 118 (2006).
- D. A. Luzzo, P. Hasper, K. A. Albert, M. A. Bibby and E. A. Martinelli Jr., Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students, *Journal of Counseling Psychology* **46**(2), 233 (1999).
- A. Martens, M. Johns, J. Greenberg and J. Schimel, Combating stereotype threat: The effect of self-affirmation on women's intellectual performance, *Journal of Experimental Social Psychology* **42**, 236 (2006).
- D. M. Marx and J. S. Roman, Female role models: Protecting women's math test performance," *Personality and Social Psychology Bulletin* **28**, 1183 (2002).
- Mastering Physics, <http://www.masteringphysics.com/>.
- T. Matsui, K. Matsui and R. Ohnishi, Mechanisms underlying math self-efficacy learning of college students, *Journal of Vocational Behavior* **37**, 225 (1990).
- E. Mazur, *Peer Instruction: A Users Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- T. L. McCaskey, M. H. Dancy and A. Elby, "Effects in assessment caused by splits between belief and understanding" in *Proceedings of the 2003 Physics Education Research Conference*, edited by J. Marx, S. Franklin and K. Cummings (AIP Conference Proceedings, New York, 2004) p. 37.
- L. McCullough, Gender, context, and physics assessment, *Journal of International Women's Studies* **5**, 20 (2004).

- L. C. McDermott and E. F. Redish, Resource Letter: PER-1: Physics Education Research, *American Journal of Physics* **67**, 755 (1999).
- L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- D. E. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible “hidden variable” in diagnostic pre-test scores, *American Journal of Physics* **70**, 1259 (2002).
- W. Mendenhall and T. Sincich, *A Second Course in Statistics: Regression Analysis* (Pearson Education, Inc. Upper Saddle River, NJ, 2003).
- K. D. Multon and S. D. Brown, Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation, *Journal of Counseling Psychology* **38**(1), 30 (1991).
- National Academy of Science, *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* (National Academy Press, Washington, DC, 2006).
- National Academy of Science, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Future* (National Academy Press, Washington, DC, 2005).
- National Research Council, *To Recruit and Advance: Women Students and Faculty in Science and Engineering* (National Academies Press, Washington DC, 2006).
- National Science Board, *The Science and Engineering Workforce: Realizing America's Potential*, NSF 03-69 (National Science Foundation, Arlington, VA, 2003).
- National Science Foundation, Division of Science Resources Statistics, *S&E Degrees: 1966 – 2006*, NSF 08-321 (Arlington, VA, 2006). Available from <http://www.nsf.gov/statistics/degrees>.
- National Science Foundation, Division of Science Resources Statistics, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2011*, NSF 11-309 (Arlington, VA, 2011). Available from <http://www.nsf.gov/statistics/women/>.
- National Science Foundation, *New Formulas for America's Workforce: Girls in Science and Engineering*, NSF 03-207 (Arlington, VA, 2003). Available from <http://www.nsf.gov/pubs/2003/nsf03207/start.htm>.
- National Science Foundation, *New Formulas for America's Workforce 2: Girls in Science and Engineering*, NSF 06-60 (Arlington, VA, 2006). Available from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf0660.

- J. Oakes, "Opportunities, achievement, and choice: Women and minority students in science and mathematics" in *Review of Research in Education*, edited by C. Cazden (The American Educational Research Association, Washington, DC, 1990) p. 153.
- J. Osborne, S. Simon and S. Collins, Attitudes towards science: A review of the literature and its implications, *International Journal of Science Education* **25**, 1049 (2003).
- V. Otero, N. D. Finkelstein, R. McCray and S. Pollock, Who is responsible for preparing science teachers? *Science* **313**, 445 (2006).
- F. Pajares, Self-efficacy beliefs in academic settings, *Review of Educational Research* **66**(4), 543 (1996).
- F. Pajares and J. Kranzler, Self-efficacy beliefs and general mental ability in mathematical problem-solving, *Contemporary Educational Psychology* **20**, 426 (1995).
- F. Pajares and M. D. Miller, Mathematics self-efficacy and mathematics performances: The need for specificity of assessment, *Journal of Counseling Psychology* **42**(2), 190 (1995).
- F. Pajares and M. D. Miller, Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis, *Journal of Educational Psychology* **86**(2), 193 (1994).
- J. E. Parsons, J. L. Meece, T. F. Adler and C. M. Kaczala, Sex differences in attributions and learned helplessness, *Sex Roles* **8**(4), 421 (1982).
- K. K. Perkins, M. M. Gratny, W. K. Adams, N. D. Finkelstein and C. E. Wieman, "Towards characterizing the relationship between students' interest in and their beliefs about physics" in *Proceedings of the 2005 Physics Education Research Conference*, edited by P. Heron, L. McCullough and J. Marx (AIP, New York, 2006), p. 137.
- K. Piatek-Jimenez, Images of mathematicians: A new perspective on the shortage of women in mathematical careers, *Mathematics Education* **40**, 633 (2008).
- G. Potvin, Z. Hazari, R. H. Tai and P. M. Sadler, Unraveling bias from student evaluations of their high school science teachers, *Science Education* **93**(5), 827 (2009).
- G. Ramirez and S. L. Beilock, Writing about testing worries boosts exam performance in the classroom, *Science* **331**, 211 (2011).
- E. F. Redish, *Teaching Physics With the Physics Suite* (John Wiley & Sons, Inc., Hoboken, NJ, 2003).
- L. A. Robertson, Why are there so few female physicists? *The Physics Teacher* **44**, 177 (2006).
- P. J. Rottinghaus, L. M. Larson and F. H. Borgen, The relation of self-efficacy and interests: A meta-analysis of 60 samples, *Journal of Vocational Behavior* **62**, 221 (2003).

- C. P. Russel and D. P. French, Factors affecting participation in traditional and inquiry-based laboratories, *Journal of College Science Teaching* **31**, 225 (2002).
- P. M. Sadler and R. H. Tai, Success in introductory college physics: the role of high school preparation, *Science Education* **85**, 111 (2001).
- V. Sawtelle, E. Brewe and L. Kramer, “Positive impacts of Modeling Instruction on self-efficacy” in *Proceedings of the 2010 Physics Education Research Conference*, AIP Conference Proceedings Number 1289, edited by C. Singh, M. Sabella and S. Rebello (AIP, New York, 2010), p. 289.
- K. Scantlebury and D. Baker, “Gender issues in science education research: Remembering where the difference lies” in *Handbook of Research on Science Education*, edited by S. Abell and N. Lederman (Lawrence Erlbaum, Mahwah, NJ, 2007) p. 257.
- L. Schiebinger, *Has Feminism Changed Science?* (Harvard University Press, Cambridge, MA, 1999).
- T. Schmader, Gender identification moderates stereotype threat effects on women’s math performance, *Journal of Experimental Social Psychology* **38**, 194 (2002).
- T. Schmader and M. Johns, Converging evidence that stereotype threat reduces working memory capacity, *Journal of Personality and Social Psychology* **85**, 440. (2003).
- T. Schmader, M. Johns and M. Barquissau, The costs of accepting gender differences: The role of stereotype endorsement in women’s experience in the math domain, *Sex Roles: A Journal of Research* **50**, 835 (2004).
- M. Schneider, Encouragement of women physics majors at Grinnell college: A case study, *The Physics Teacher* **39**, 280 (2001).
- E. Seymour, Guest comment: Why undergraduates leave the sciences, *American Journal of Physics* **63**(3), 199 (1995).
- E. Seymour and N. M. Hewitt, *Talking About Leaving: Why Undergraduates Leave the Sciences* (Westview Press, Boulder, CO, 1997).
- A. Sfard and A. Prusak, Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity, *Educational Researcher* **34**(4), 14 (2005).
- K. A. Shaw, “The development of a physics self-efficacy instrument for use in the introductory classroom,” in *Proceedings of the 2003 Physics Education Research Conference*, AIP Conference Proceedings Number 720, edited by J. Marx, S. Franklin and K. Cummings (AIP, New York, 2003), p. 137.

- D. K. Sherman and G. L. Cohen, "The psychology of self-defense: Self-affirmation theory" in *Advances in Experimental Social Psychology*, edited by M. P. Zanna (Academic Press, San Diego, CA, 2006), p. 183.
- M. Shih, T. L. Pittinsky and N. Ambady, Stereotype susceptibility: Identity salience and shifts in quantitative performance, *Psychological Science* **10**, 80 (1999).
- R. M. Sirkin, *Statistics for the Social Sciences* (SAGE Publications, Thousand Oaks, CA, 1999).
- S. J. Spencer, C. M. Steele and D. M. Quinn, Stereotype threat and women's math performance, *Journal of Experimental Social Psychology* **34**, 4 (1999).
- F. K. Stage and S. A. Maple, Incompatible Goals: Narratives of graduate women in the mathematics pipeline, *American Educational Research Journal* **33**(1), 23 (1996).
- W. B. Stanley and N. W. Brickhouse, Multiculturalism, universalism, and science education, *Science Education* **78**(4), 387 (1994).
- C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance, *American Psychologist* **52**, 613 (1997).
- C. M. Steele, "The psychology of self-affirmation: Sustaining the integrity of the self," in *Advances in Experimental Social Psychology*, edited by L. Berkowitz (Academic Press, New York, 1988) p. 261.
- C. M. Steele and J. Aronson, Stereotype threat and the intellectual test performance of African Americans, *Journal of Personality and Social Psychology* **69**, 797 (1995).
- C. M. Steele, S. J. Spencer and J. Aronson, "Contending with group image: The psychology of stereotype and social identity threat" in *Advances in Experimental Social Psychology* **34**, 379 (2002).
- J. Stevens, *Applied Multivariate Statistics for the Social Sciences* (Erlbaum, Mahwah, NJ, 2002).
- J. Stone, C. I. Lynch, M. Sjomeling and J. M. Darley, Stereotype threat effects on Black and White athletic performance, *Journal of Personality and Social Psychology* **77**, 1213 (1999).
- J. Stone and C. McWhinnie, Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes, *Journal of Experimental Social Psychology* **44**, 445 (2008).
- J. G. Stout, N. Dasgupta, M. Husinger and M. A. McManus, STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM), *Journal of Personality and Social Psychology* **100**(2), 255 (2011).

- G. Taasoobshirazi and M. Carr, Gender differences in science: An expertise perspective, *Educational Psychology Review* **20**, 149 (2008).
- R. H. Tai, and P. M. Sadler, Gender differences in introductory undergraduate physics performance: university physics versus college physics in the USA, *International Journal of Science Education* **23**, 1017 (2001).
- B. D. Tatum, *Why are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race* (Basic Books, New York, 1997).
- Teaching the Majority: Breaking the Gender Barrier in Science, Mathematics, and Engineering*, edited by S. V. Rosser (Teacher's College Press, New York, 1995).
- B. Thomadsen, The shortage of radiotherapy physicists, *Journal of the American College of Radiology* **1**(4), 280 (2004).
- R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation, *American Journal of Physics* **66**, 338 (1998).
- K. Tonso, Student engineers and engineer identity: Campus engineer identities as figured world, *Cultural Studies of Science Education* **1**, 273 (2006).
- C. Turpen and N. D. Finkelstein, Not all interactive engagement is the same: Variations in physics professors' implementation of *Peer Instruction*, *Physical Review Special Topics – Physics Education Research* **5**, 020101 (2009).
- E. L. Usher and F. Pajares, Inviting confidence in school: Invitations as a critical source of the academic self-efficacy beliefs of entering middle school students, *Journal of Invitational Theory and Practice* **12**, 7 (2006).
- E. L. Usher and F. Pajares, Sources of academic and self-regulatory efficacy beliefs of entering middle school students, *Contemporary Educational Psychology* **31**, 125 (2006).
- E. L. Usher and F. Pajares, Sources of self-efficacy in school: Critical review of the literature and future directions, *Review of Educational Research* **78**(4), 751 (2008).
- V. Valian, *Why So Slow: The Advancement of Women* (The MIT Press, Cambridge, 1999).
- G. M. Walton and G. L. Cohen, A question of belonging: Race, social fit, and achievement, *Journal of Personality and Social Psychology* **92**(1), 82 (2007).
- M. Weinburgh, Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991, *Journal of Research in Science Teaching* **32**, 387 (1995).
- E. Wenger, *Communities of Practice: Learning, Meaning, and Identity* (Cambridge University Press, Cambridge, 1998).

- S. White and C. L. Tesfaye, *Who teaches high school physics? Results from the 2008-09 nationwide survey of high school physics teachers* (AIP Statistical Research Center, College Park, MD, 2010). Available from www.aip.org/statistics
- B. L. Whitten, S. R. Foster, M. L. Duncombe, P. E. Allen, P. Heron, H. M. Zorn, L. McCullough, K. A. Shaw and B. A. P. Taylor, What works? Increasing the participation of women in undergraduate physics, *Journal of Women and Minorities in Science and Engineering* **9**, 239 (2003).
- S. D. Willoughby and A. Metz, Exploring gender differences with different gain calculations in astronomy and biology, *American Journal of Physics* **77**, 651 (2009).
- N. C. Yeung and C. von Hippel, Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers, *Accident Analysis and Prevention* **40**, 667 (2008).
- V. Y. Yzerbyt, D. Muller and C. M. Judd. Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions, *Journal of Experimental Social Psychology* **40**, 424 (2004).
- A. L. Zeldin, S. L. Britner and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, *Journal of Research in Science Teaching* **45**(9), 1036 (2008).
- A. L. Zeldin and F. Pajares, Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers, *American Educational Research Journal* **37**(1), 215 (2000).
- B. J. Zimmerman, Self-efficacy: An essential motive to learn, *Contemporary Educational Psychology* **25**, 82 (2000).

CHAPTER 2 APPENDIX

Updated Results for Entire Data Set

Data on FMCE pre- and post-test gender gaps have been collected between Spring 2004 and Fall 2010. Though only some of the data were used for the study presented in this chapter, in this appendix, we present the results from all semesters between Spring 2004 and Fall 2009 (in the Spring and Fall 2010 semesters we conducted an intervention study, discussed in Chapter 6, so we do not present results from those semesters here). Over the 12 semesters, we have FMCE data from 4028 students (2970 males and 1058 females), and course performance data from 6372 students (4839 males and 1553 females). Figure and Table numbers correspond to matching figures and tables in the chapter.

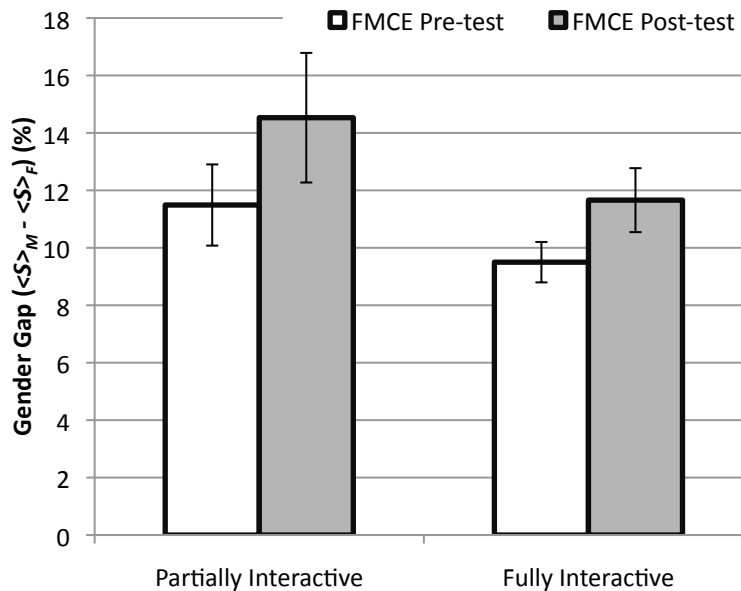


Figure 1A. Pre- and post-test gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) for partially and fully interactive courses. Student performance on the FMCE is averaged over three semesters (partially interactive) and nine semesters (fully interactive). Error bars represent the standard errors of the mean. The pre-test [$t(4027)=1.26, p=0.21$] and post-test [$t(4027)=1.14, p=0.25$] gender gaps are not significantly different between partially and fully interactive courses. There is no statistically significant shift in the gender gap from pre- to post-test for either partially [$t(788)=1.14, p=0.25$] or fully [$t(3238)=1.64, p=0.10$] interactive courses.

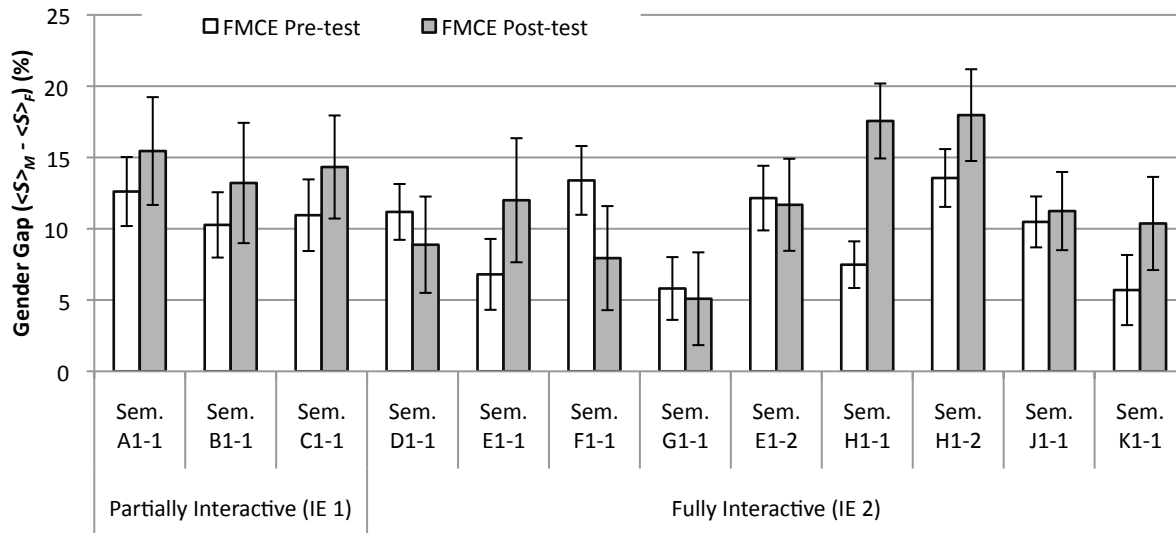


Figure 2A. Gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) in each semester. The data represent ten different instructors and over 4000 students. Error bars represent the standard errors of the mean.

Table 3A. Analysis of students' course grades. Each column contains the difference between the average scores for males and females ($\langle S \rangle_M - \langle S \rangle_F$). Error (shown in parentheses) is computed from the standard errors of the mean for males and females added in quadrature. The * indicates the gender difference is statistically significant at the $p < 0.05$ level (via two-tailed t -test).

	Participation (%)	Homework (%)	Exams (%)	Course GPA (4 pt scale)
Semester A1-1		-5.0 (2.0) *	3.4 (1.2) *	0.11 (0.11)
Semester B1-1		-4.8 (1.8) *	3.7 (1.6) *	0.10 (0.11)
Semester C1-1		-5.0 (2.0) *	6.3 (1.5) *	0.10 (0.10)
Semester D1-1	-6.6 (1.6) *	-7.6 (1.9) *	4.9 (1.6) *	0.04 (0.10)
Semester E1-1	-4.9 (1.8) *	-2.9 (1.9)	5.2 (1.5) *	0.17 (0.11)
Semester F1-1	-8.1 (1.8) *	-2.0 (2.0)	4.8 (1.6) *	0.15 (0.12)
Semester G1-1	-3.0 (1.6)	-3.0 (2.0)	3.3 (1.4) *	0.06 (0.11)
Semester E1-2	-5.2 (2.1) *	-5.2 (2.2) *	2.5 (1.8)	-0.01 (0.10)
Semester H1-1	-4.2 (1.6) *	-3.0 (1.8)	5.5 (1.4) *	0.15 (0.10)
Semester H1-2	-3.8 (1.7) *	-2.9 (1.9)	7.2 (1.6) *	0.28 (0.10) *
Semester J1-1	-6.4 (1.5) *	-4.2 (1.5) *	5.2 (1.3) *	0.17 (0.09) *
Semester K1-1	-5.3 (2.5) *		4.9 (1.6) *	0.06 (0.10)
Average	-5.3 (0.7) *	-4.3 (0.6) *	4.9 (0.5) *	0.12 (0.03) *

CHAPTER 3 APPENDIX

Updated Results for Entire Data Set

Data on gender gaps have been collected between Spring 2004 and Spring 2010. Though only some of the data were used for the study presented in this chapter, in this appendix, we present the results from all semesters between Fall 2004 and Spring 2010. Over the 12 semesters, we have BEMA data from 3310 students (2451 males and 859 females), and course performance data from 4832 students (3654 males and 1178 females). Figure and Table numbers correspond to matching figures and tables in the chapter.

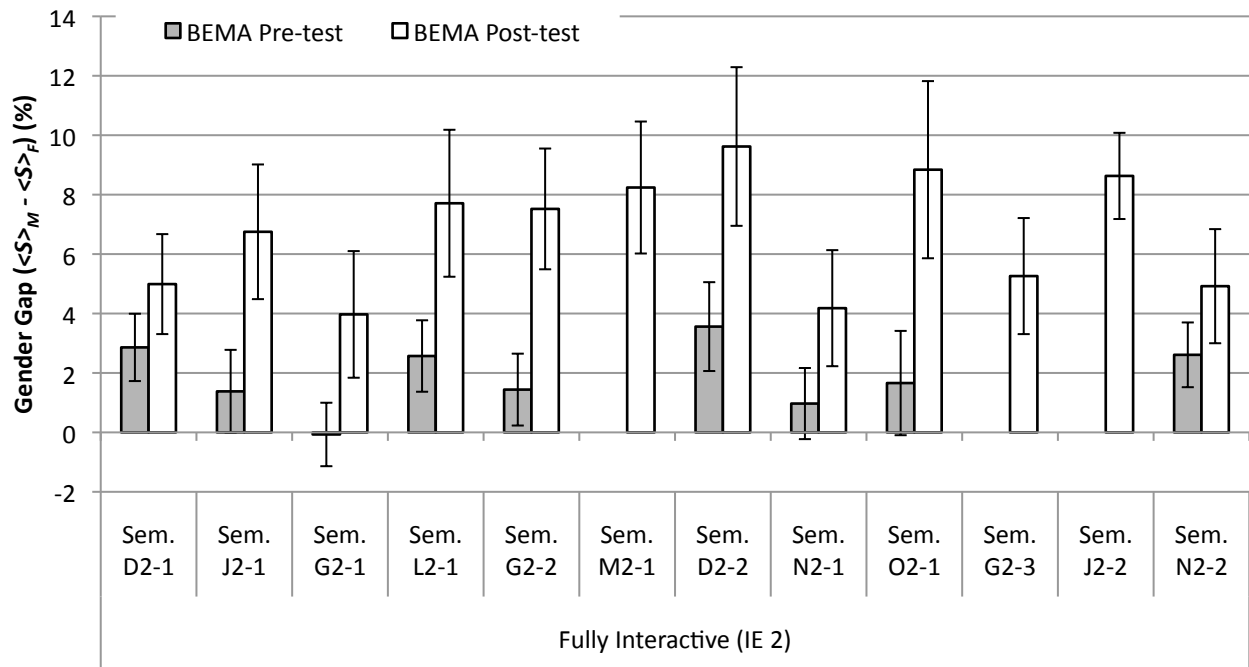


Figure 2A. Pre- and post-test gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) by semester. The data shown here include all students who took the pre- and post-BEMA. These data represent seven different instructors and over 3000 students. The semester is indicated along the x-axis. The error bars represent the standard errors of the mean.

Table 5A. Analysis of students' course grades. Each column contains the difference between the average scores for males and females ($\langle S \rangle_M - \langle S \rangle_F$). Error (shown in parentheses) is computed from the standard errors of the mean for males and females added in quadrature. The asterisk (*) indicates that the difference is statically significant at the $p < 0.05$ level. In semester L2-1 no participation credit was given.

	Participation (%)	Homework (%)	Exams (%)	Course GPA (4 pt. scale)
Semester D2-1	-6.4 (1.0)*	-7.0 (1.5)*	3.6 (1.6)*	-0.04 (0.12)
Semester J2-1	-6 (3)*	-4 (2)	6.1 (1.9)*	0.08 (0.13)
Semester G2-1	-7.4 (1.6)*	-3.3 (1.8)	4.2 (1.5)*	-0.03 (0.12)
Semester L2-1		-3 (3)	2.5 (1.7)	0.06 (0.13)
Semester G2-2	-7.2 (1.6)*	-5.5 (1.9)*	4.8 (1.7)*	0.07 (0.12)
Semester M2-1	-5 (2)*	-4.0 (1.7)*	2.4 (1.6)	-0.01 (0.13)
Semester D2-2	-4.9 (1.7)*	-2.6 (1.9)	3.1 (1.4)*	0.02 (0.11)
Semester N2-1	-11.5 (1.8)*	-8.2 (1.9)*	-1.5 (1.6)	-0.32 (0.11)*
Semester O2-1	-2.7 (1.9)	-3.7 (1.7)*	7.7 (1.4)*	0.20 (0.09)*
Semester G2-3	-6.1 (1.7)*	-5.3 (1.8)*	4.2 (1.4)*	0.02 (0.11)
Semester J2-2	-8.2 (1.9)*	-2.1 (1.8)	7.7 (1.5)*	0.35 (0.09)*
Semester N2-2	-6.7 (2.3)*	-3.6 (2.3)	2.7 (1.7)	0.04 (0.13)
Average	-6.4 (0.6)*	-4.3 (0.6)*	4.1 (0.5)*	0.04 (0.03)

CHAPTER 4 APPENDICES

Appendix A: Physics Self-Efficacy and Identity Survey

We include both the pre- and post-course surveys. Comments in square brackets [] indicate how the question was asked in the K1-1 and J2-2 semesters. Comments in the curly brackets { } indicate how the question was asked for students in Physics 2. Items with an * were not asked in the K1-2 and G2-4 semesters.

Pre-Course Survey:

- 1) Do you see yourself as a science person? Why or why not? *
- 2) Do you see yourself as a *physics* person? Why or why not? *
- 3) What are you planning to major in? What made you decide to choose that major? Are there any specific experiences that you recall that helped you decide? *
- 4) Did you ever consider being a physics major? What made you decide to choose or not choose physics? *

Please rate the extent to which you agree or disagree with each of the following statements:

- 5) Other students recognize me as a physics person.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 6) A teacher encouraged me to pursue physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 7) My parents have encouraged me to pursue physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 8) I see myself as a physics person.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 9) I feel like I could be a good physicist.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

10) A teacher has told me that I'm good at physics.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

Please rate the extent to which you agree or disagree with the following statements about your *experiences in prior physics [science] courses*:

11) I was unable to think clearly when taking exams.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

12) Physics [Science] makes me feel uneasy.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

13) I got really stressed when working on homework.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

14) I worried about my ability to solve problems on exams.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

15) I enjoyed my prior physics [science] classes.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

16) Please rate how *confident* you are that you can do each of the following things *in this physics course*:

	Not Confident	A Little Confident	Somewhat Confident	Mostly Confident	Very Confident	Not Answered
Complete homework assignments by myself						
Complete homework assignments with the help of other students						
Perform well on exams						
Demonstrate what I know on exams						
Discuss physics ideas with my peers						

	Not Confident	A Little Confident	Somewhat Confident	Mostly Confident	Very Confident	Not Answered
Defend my physics ideas to my peers						
Find other students to help me when I get stuck on homework *						
Ask the professor for help when I get stuck on homework *						
Ask a TA for help when I get stuck on homework *						
Learn physics concepts						
Work well in a group during recitation *						
Find a group of students to study with *						
Express my opinions when others disagree with me						
Discuss physics ideas when I don't understand them						
Manipulate algebraic equations						
Use calculus to solve physics problems						
Interpret graphs						
Pay attention in lecture *						
Complete the course with a B or better						
Answer questions posed by the professor in class						
Ask questions during lecture *						

17) How confident are you that you will *earn a B or better* in Physics 1 {2}? Please explain your response. *

18) How confident are you that you will *learn the physics concepts* in Physics 1 {2}? Please explain your response. *

19) Please report your gender. *

Female

Male

Prefer not to answer

20)

	Yes	No	Not Answered
Did you take a <i>regular</i> physics class in high school? *			
Did you take an <i>honors</i> physics class in high school? *			
Did you take an <i>AP</i> physics class in high school? *			
Did you take any other kind of physics class in high school? *			

21) What was the grade you received in your *last high school physics* course? *

A B C D F I did not take a physics class in high school.

22) Why are you taking Physics 1 {2}? (Because it is required for you major? Because you are interested in physics? Because you are thinking about majoring in physics? To fulfill a general science requirement? Some other reason?) *

Post-Course Survey:

- 1) Do you see yourself as a *physics* person? Why or why not? *
- 2) Has your answer to the above question changed since the beginning of the semester? Why or why not? *
- 3) Do you feel like you could be a good physicist? Why or why not? *

Please rate the extent to which you agree or disagree with each of the following statements:

4) Other students recognize me as a physics person.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
----------------------	----------------------	---------	-------------------	-------------------

5) A teacher encouraged me to pursue physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
----------------------	----------------------	---------	-------------------	-------------------

6) My parents have encouraged me to pursue physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
----------------------	----------------------	---------	-------------------	-------------------

7) I see myself as a physics person.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
----------------------	----------------------	---------	-------------------	-------------------

8) I feel like I could be a good physicist.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
----------------------	----------------------	---------	-------------------	-------------------

9) A teacher has told me that I'm good at physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
----------------------	----------------------	---------	-------------------	-------------------

Please rate the extent to which you agree or disagree with the following statements about your *experiences in this physics course*:

10) I was unable to think clearly when taking exams in this course.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

11) Physics makes me feel uneasy.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

12) I got really stressed when working on homework in this class.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

13) I worried about my ability to solve problems on exams in this class.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

14) I enjoyed physics this semester.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

15) Please rate how *confident* you are that you can do each of the following things *in Physics 2 {your next physics course}*:

Note: If you **ARE NOT** planning to take Physics 2 {another physics course}, please rate how *confident* you are that you can do each of the following things *in your next science course*.

	Not Confident	A Little Confident	Somewhat Confident	Mostly Confident	Very Confident	Not Answered
Complete homework assignments by myself						
Complete homework assignments with the help of other students						
Perform well on exams						
Demonstrate what I know on exams						
Discuss physics ideas with my peers						
Defend my physics ideas to my peers						
Find other students to help me when I get stuck on homework *						

	Not Confident	A Little Confident	Somewhat Confident	Mostly Confident	Very Confident	Not Answered
Ask the professor for help when I get stuck on homework *						
Ask a TA for help when I get stuck on homework *						
Learn physics concepts						
Work well in a group during recitation *						
Find a group of students to study with *						
Express my opinions when others disagree with me						
Discuss physics ideas when I don't understand them						
Manipulate algebraic equations						
Use calculus to solve physics problems						
Interpret graphs						
Pay attention in lecture *						
Complete the course with a B or better						
Answer questions posed by the professor in class						
Ask questions during lecture *						

16) What experiences did you have in this course that contributed to how confident you are about succeeding in Physics 2 {your next physics course}? *

Please rate the extent to which you agree or disagree with the following statements about your *experiences in this physics course*:

17) I received good grades on my homework in this class.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

18) Watching other students in class made me think that I could succeed in physics.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

19) My peers in this class encouraged me to do well on homework.

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

20) I was usually able to help my classmates with difficult physics problems.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

21) Listening to other students during Tutorials made me think that I could *not* understand physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

22) The instructor in this course encouraged me to put forth my best effort.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

23) I rarely knew the answer to the clicker questions raised in class.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

24) Students who were similar to me did well on exams.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

25) No one in this class has encouraged me to go on in science after this course.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

26) People often tell me that I'm good at physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

27) I had difficulty with exams in this class.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

28) Please report your gender. *

Female	Male	Prefer not to answer
--------	------	----------------------

Appendix B: Factor Analysis Results

In the tables below we give the rotated factor loadings for males and females on the pre- and post-course survey. Only those factor loadings that are greater than 0.33 are given. The highest factor loading for each task is boldfaced.

Factor loadings for items from the pre-course survey, for male students only.

	Factor 1	Factor 2	Factor 3	Factor 4
Complete homework assignments by myself	0.65	0.39		
Perform well on exams	0.70	0.41		
Demonstrate what I know on exams	0.63	0.45		
Discuss physics ideas with my peers		0.74		
Defend my physics ideas to my peers	0.33	0.77		
Learn physics concepts	0.61	0.39		
Complete the course with a B or better	0.66	0.36		
Answer questions posed by the professor in class	0.42	0.56		0.33
Manipulate algebraic equations	0.71			
Use calculus to solve physics problems	0.75			
Interpret graphs	0.72			
Complete homework assignments with the help of other students	0.38		0.51	
Find other students to help me when I get stuck on homework			0.84	
Find a group of students to study with			0.81	
Work well in a group during recitation		0.41	0.49	
Ask the professor for help when I get stuck on homework				0.79
Ask a TA for help when I get stuck on homework				0.78
Ask questions during lecture		0.54		0.46
Express my opinions when others disagree with me		0.65		
Discuss physics ideas when I don't understand them		0.55	0.36	
Pay attention in lecture				0.63

Factor loadings for items from the pre-course survey, for female students only.

	Factor 1	Factor 2	Factor 3	Factor 4
Complete homework assignments by myself	0.69			
Perform well on exams	0.86			
Demonstrate what I know on exams	0.81			
Discuss physics ideas with my peers	0.68	0.46		
Defend my physics ideas to my peers	0.69	0.39		
Learn physics concepts	0.71			
Complete the course with a B or better	0.76			
Answer questions posed by the professor in class	0.64			0.37
Manipulate algebraic equations			0.84	
Use calculus to solve physics problems			0.82	
Interpret graphs			0.78	
Complete homework assignments with the help of other students	0.44	0.53		
Find other students to help me when I get stuck on homework		0.83		
Find a group of students to study with		0.85		
Work well in a group during recitation		0.52		
Ask the professor for help when I get stuck on homework				0.84
Ask a TA for help when I get stuck on homework				0.83
Ask questions during lecture	0.34			0.57
Express my opinions when others disagree with me	0.43	0.63		
Discuss physics ideas when I don't understand them		0.52		0.34
Pay attention in lecture			0.40	

Factor loadings for items from the post-course survey, for male students only.

	Factor 1	Factor 2	Factor 3	Factor 4
Complete homework assignments by myself	0.73			
Perform well on exams	0.85			
Demonstrate what I know on exams	0.83			
Discuss physics ideas with my peers	0.66			
Defend my physics ideas to my peers	0.73			
Learn physics concepts	0.67			
Complete the course with a B or better	0.76			
Answer questions posed by the professor in class	0.58			0.51
Manipulate algebraic equations			0.83	
Use calculus to solve physics problems			0.80	
Interpret graphs	0.33		0.74	
Complete homework assignments with the help of other students	0.38	0.55		
Find other students to help me when I get stuck on homework		0.88		
Find a group of students to study with		0.84		
Work well in a group during recitation		0.58		
Ask the professor for help when I get stuck on homework		0.61		0.40
Ask a TA for help when I get stuck on homework		0.63		0.36
Ask questions during lecture				0.65
Express my opinions when others disagree with me	0.46	0.39		0.37
Discuss physics ideas when I don't understand them		0.35		0.41
Pay attention in lecture				0.76

Factor loadings for items from the post-course survey, for female students only.

	Factor 1	Factor 2	Factor 3	Factor 4
Complete homework assignments by myself	0.62			
Perform well on exams	0.87			
Demonstrate what I know on exams	0.85			
Discuss physics ideas with my peers	0.58	0.45		
Defend my physics ideas to my peers	0.64	0.35		
Learn physics concepts	0.71			
Complete the course with a B or better	0.76			
Answer questions posed by the professor in class	0.60			0.35
Manipulate algebraic equations			0.81	
Use calculus to solve physics problems			0.80	
Interpret graphs			0.77	
Complete homework assignments with the help of other students		0.67		
Find other students to help me when I get stuck on homework		0.86		
Find a group of students to study with		0.84		
Work well in a group during recitation		0.60		0.36
Ask the professor for help when I get stuck on homework				0.80
Ask a TA for help when I get stuck on homework				0.77
Ask questions during lecture	0.46			0.49
Express my opinions when others disagree with me	0.50	0.53		
Discuss physics ideas when I don't understand them	0.39	0.50		0.34
Pay attention in lecture				0.56

Appendix C: Summary Tables of Survey Statistics

This appendix contains summary tables of the average values of self-efficacy (and each of its components) and the sources of self-efficacy for males and females in each semester of Physics 1 and Physics 2, and averaged over all semesters of each course. The * (**) indicates that the difference between males and females (for all values in the M – F columns) or between the pre- and post-surveys (in the POST – PRE rows) are statistically significant at the $p < 0.05$ ($p < 0.01$) level. The final two columns in each table contain p values for ANOVA tests to determine whether there are significant differences in the average value for males or females by semester. Throughout all tables, significant p values are underlined.

Physics Self-Efficacy

Physics I:

	Semester K1-1			Semester J1-2			Semester K1-2			All Semesters			ANOVA <i>p</i> value	
	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F	M	F
PRE														
Total	3.95	3.86	0.09	4.03	3.74	0.29**				4.00	3.79	0.21**	0.14	0.19
Performance	4.00	3.68	0.32**	4.03	3.57	0.46**	3.73	3.20	0.53**	3.94	3.51	0.43**	<0.01	<0.01
Math	4.19	4.19	0.00	4.29	4.08	0.21**	4.01	3.93	0.08	4.18	4.08	0.10	<0.01	0.16
Getting Help	3.64	3.78	-0.14	3.72	3.56	0.16				3.69	3.65	0.04	0.27	0.06
Work w/ Others	3.99	4.05	-0.06	4.18	4.01	0.17*				4.10	4.03	0.07	<0.01	0.69
POST														
Total	3.66	3.56	0.10	3.81	3.45	0.36**				3.75	3.50	0.25**	0.01	0.27
Performance	3.62	3.26	0.36**	3.66	3.07	0.59**	3.31	2.71	0.60**	3.54	3.02	0.52**	<0.01	<0.01
Math	4.08	4.03	0.05	4.21	3.97	0.24**	3.96	3.97	-0.01	4.10	3.99	0.11	<0.01	0.87
Getting Help	3.26	3.36	-0.10	3.51	3.30	0.21*				3.41	3.33	0.08	<0.01	0.63
Work w/ Others	3.89	3.99	-0.10	4.10	3.92	0.18*				4.02	3.95	0.07	<0.01	0.58
POST - PRE														
Total	-0.29**	-0.30**	0.01	-0.22**	-0.29**	0.07				-0.25**	-0.29**	0.04	0.12	0.93
Performance	-0.38**	-0.42**	0.04	-0.37**	-0.50**	0.13	-0.42**	-0.49**	0.07	-0.40**	-0.49**	0.09	0.34	0.80
Math	-0.11*	-0.16*	0.05	-0.08	-0.11	0.03	-0.05	0.04	-0.09	-0.08**	-0.09*	0.01	0.75	0.47
Getting Help	-0.38**	-0.42**	0.04	-0.21**	-0.26**	0.05				-0.28**	-0.32**	0.04	0.01	0.19
Work w/ Others	-0.10	-0.06	-0.04	-0.08*	-0.09	0.01				-0.08**	-0.08	0.00	0.84	0.85

Physics 2:

	ANOVA																	
	Semester J2-2				Semester N2-2				Semester G2-4				All Semesters					
	M	F	M-F	M	M	F	M-F	M	M	F	M-F	M	M	F	M-F	M	F	
PRE																		
Total	3.93	3.67	0.26**	3.90	3.80	3.80	0.10	3.79	3.24	3.72	0.19**	3.91	3.72	3.72	0.19**	3.91	3.72	0.21
Performance	3.90	3.32	0.58**	3.93	3.57	3.57	0.36**	3.79	3.24	3.36	0.51**	3.87	3.36	3.36	0.51**	3.87	3.36	0.11
Math	4.34	4.29	0.05	4.29	4.29	4.29	0.00	4.17	4.04	4.21	0.03	4.24	4.21	4.21	0.03	4.24	4.21	0.08
Getting Help	3.55	3.51	0.04	3.51	3.71	3.71	-0.20			3.59	-0.02	3.57	3.59	3.59	-0.02	3.57	3.59	0.11
Work w/ Others	4.10	4.09	0.01	4.09	3.93	3.93	0.16			4.03	-0.02	4.01	4.03	4.03	-0.02	4.01	4.03	<0.01
																		0.19
POST																		
Total	3.84	3.39	0.45**	3.80	3.66	3.66	0.14	3.73	3.20	3.50	0.32**	3.82	3.50	3.50	0.32**	3.82	3.50	0.01
Performance	3.78	3.02	0.76**	3.80	3.41	3.41	0.39**	3.73	3.20	3.18	0.53**	3.77	3.18	3.18	0.53**	3.77	3.18	0.62
Math	4.27	3.98	0.29**	4.23	4.21	4.21	0.02	4.16	4.08	4.07	0.15*	4.22	4.07	4.07	0.15*	4.22	4.07	0.20
Getting Help	3.47	3.15	0.32**	3.37	3.43	3.43	-0.06			3.26	0.16	3.42	3.26	3.26	0.16	3.42	3.26	0.24
Work w/ Others	4.04	3.95	0.09	3.92	4.00	4.00	-0.08			3.97	0.01	3.98	3.97	3.97	0.01	3.98	3.97	0.73
POST - PRE																		
Total	-0.09*	-0.28**	0.19**	-0.10**	-0.14	-0.14	0.04	-0.06	-0.04	-0.22**	0.13**	-0.09**	-0.22**	-0.22**	0.13**	-0.09**	-0.22**	0.12
Performance	-0.12**	-0.30**	0.18*	-0.13**	-0.16	-0.16	0.03	-0.06	-0.04	-0.18**	0.08	-0.10**	-0.18**	-0.18**	0.08	-0.10**	-0.18**	0.05
Math	-0.07	-0.31**	0.24**	-0.06	-0.08	-0.08	0.02	-0.01	0.04	-0.14*	0.12	-0.02	-0.14*	-0.14*	0.12	-0.02	-0.14*	0.45
Getting Help	-0.08	-0.36**	0.28**	-0.14**	-0.28**	-0.28**	0.14			-0.33**	0.18*	-0.15**	-0.33**	-0.33**	0.18*	-0.15**	-0.33**	0.64
Work w/ Others	-0.06	-0.14	0.08	-0.17	0.07	0.07	-0.24			-0.06	0.03	-0.03	-0.06	-0.06	0.03	-0.03	-0.06	0.13

Sources of Self-Efficacy

Physics 1:

	Semester K1-1			Semester J1-2			Semester K1-2			All Semesters			ANOVA p value	
	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F	M	F
PRE														
Emotional Response	3.74	3.52	0.22*	3.66	3.23	0.43**	3.33	2.76	0.57**	3.60	3.21	0.39**	<0.01	<0.01
POST														
Mastery Experiences	3.49	3.21	0.28**	3.45	3.06	0.39**	3.25	2.91	0.34**	3.40	3.06	0.34**	<0.01	0.02
Vicarious Experiences	3.55	3.34	0.21*	3.59	3.23	0.36**	3.37	2.98	0.39**	3.51	3.19	0.32**	<0.01	<0.01
Verbal Persuasion	3.31	3.32	-0.01	3.45	3.38	0.07	3.22	2.94	0.28**	3.34	3.23	0.11*	<0.01	<0.01
Emotional Response	3.08	2.69	0.39**	3.13	2.43	0.70**	2.73	2.04	0.69**	2.99	2.39	0.60**	<0.01	<0.01
POST – PRE														
Emotional Response	0.66**	0.83**	0.17	0.53**	-0.80**	0.27*	-0.60**	-0.72**	0.12	-0.61**	-0.82**	0.22**	0.21	0.82

Physics 2:

	Semester J2-2			Semester N2-2			Semester G2-4			All Semesters			ANOVA p value	
	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F	M	F
PRE														
Emotional Response	3.55	3.22	0.33**	3.64	3.33	0.31*	3.26	2.67	0.59**	3.48	3.08	0.40**	<0.01	<0.01
POST														
Mastery Experiences	3.64	3.08	0.56**	3.64	3.33	0.31**	3.54	3.22	0.32**	3.60	3.19	0.41**	0.25	0.06
Vicarious Experiences	3.69	3.24	0.45**	3.66	3.38	0.28**	3.71	3.50	0.21*	3.69	3.36	0.33**	0.76	0.07
Verbal Persuasion	3.59	3.29	0.30**	3.37	3.36	0.01	3.53	3.49	0.04	3.50	3.37	0.13**	<0.01	0.13
Emotional Response	3.43	2.47	0.96**	3.35	2.80	0.55**	3.31	2.72	0.59**	3.37	2.63	0.74**	0.34	0.05
POST – PRE														
Emotional Response	-0.12*	-0.75**	0.63**	-0.29**	-0.53**	0.24	0.05	0.05	0.00	-0.11**	-0.45**	0.34**	<0.01	<0.01

Appendix D: Regression Models for Physics 2

Coefficient estimates and multiple regression model statistics for each regression model. The dependent variable is the average exam score. B is the un-standardized coefficient, β is the standardized coefficient.

	Model 1			Model 2			Model 3			Model 4		
Model-Level Statistics												
Multiple R^2	0.050			0.218			0.276			0.285		
F statistic p value	<0.001			<0.001			<0.001			<0.001		
Std. Error	13.1			11.9			11.5			11.4		
Independent Variables												
	B	β	Sig.	B	β	Sig.	B	β	Sig.	B	β	Sig.
Constant	73.3		<0.01	71.5		<0.01	54.1		<0.01	55.5		<0.01
<i>Female</i>	-6.81	-0.22	<0.01	-5.49	-0.18	<0.01	-3.53	-0.12	<0.01	-6.84	-0.22	<0.01
Math Score				5.42	0.39	<0.01	4.15	0.30	<0.01	4.17	0.30	<0.01
Semester N2-2				4.24	0.15	<0.01	3.84	0.13	<0.01	2.04	0.07	0.06
Semester G2-4				0.86	0.03	0.36	1.79	0.06	0.05	0.56	0.02	0.60
Pre-Performance SE							3.32	0.20	<0.01	3.16	0.19	<0.01
Pre-Emotional Response							1.25	0.09	0.02	1.33	0.09	0.02
Gender x Sem_N2-2										6.69	0.13	<0.01
Gender x Sem_G2-4										4.25	0.09	0.04

Coefficient estimates and multiple regression model statistics for each regression model. The dependent variable is the post-course *performance* self-efficacy score. *B* is the un-standardized coefficient, β is the standardized coefficient.

	Model 1			Model 2			Model 3			Model 4		
Model-Level Statistics												
Multiple R^2	0.084			0.344			0.551			0.700		
<i>F</i> statistic <i>p</i> value	<0.001			<0.001			<0.001			<0.001		
Std. Error	0.85			0.72			0.60			0.49		
Independent Variables	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.
Constant	3.76		<0.01	1.25		<0.01	-0.02		0.91	-0.43		<0.01
Gender	-0.59	-0.29	<0.01	-0.35	-0.18	<0.01	-0.15	-0.08	<0.01	-0.05	-0.02	0.25
Average Exam Score				0.04	0.52	<0.01	0.02	0.34	<0.01	0.01	0.12	<0.01
Semester N2-2				-0.03	-0.02	0.60	-0.02	-0.01	0.64	0.05	0.03	0.22
Semester G2-4				-0.05	-0.02	0.42	0.02	0.01	0.61	0.02	0.01	0.65
Pre-Performance SE							0.55	0.51	<0.01	0.31	0.29	<0.01
Mastery Experiences										0.23	0.20	<0.01
Vicarious Experiences										0.11	0.09	<0.01
Verbal Persuasion										0.11	0.08	<0.01
Emotional Response										0.24	0.27	<0.01

Once all of the variables in Model 4 are included, students' Math Score is not a significant predictor of post-course *performance* self-efficacy [$\beta=0.004$, $t(932)=0.23$, $p=0.82$]. All interactions between gender and each of the sources of self-efficacy were also tested. None of these interactions was significant [ME: $\beta=-0.10$, $t(929)=-1.25$, $p=0.21$; VE: $\beta=-0.06$, $t(929)=-0.95$, $p=0.34$; VP: $\beta=0.03$, $t(929)=0.47$, $p=0.64$; ER: $\beta=0.09$, $t(929)=1.60$, $p=0.11$], suggesting that there are no differences in how the sources predict self-efficacy for males and females.

CHAPTER 5 APPENDICES

Appendix A: Summary Tables of Survey Statistics

This appendix contains summary tables of the average values of physics identity for males and females in each semester of Physics 1 and Physics 2, and averaged over all semesters of each course. The * indicates that the difference between males and females (for all values in the M – F columns) or between the pre- and post-surveys (in the POST – PRE rows) are statistically significant at the $p < 0.05$ level. In the tables that are broken down by semester, the final two columns contain p values for ANOVA tests to determine whether there are significant differences in the average value for males or females by semester. Throughout all tables, significant p values are underlined.

Throughout the analysis, the Recognition by Others measure is an average of all 4 statements for the K1-1, J1-2, J2-2 and N2-2 semesters, but averages only the 2 statements that were asked in the K1-2 and G2-4 semesters.

Average Self-Identity and Recognition by Others:

Physics 1:

	Semester K1-1			Semester J1-2			Semester K1-2			All Semesters			ANOVA <i>p</i> value		
	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F
	PRE														
Self	3.79	3.33	0.46*	3.70	3.21	0.49*	3.52	2.89	0.63*	3.68	3.16	0.52*	<0.01	0.02	
Others	3.41	3.02	0.39*	3.35	3.18	0.17	3.42	2.86	0.56*	3.39	3.05	0.34*	0.56	0.08	
POST															
Self	3.48	2.93	0.55*	3.46	2.83	0.63*	3.31	2.30	1.01*	3.42	2.70	0.72*	0.15	<0.01	
Others	3.36	3.01	0.35*	3.28	3.08	0.20*	3.19	2.47	0.72*	3.28	2.88	0.40*	0.14	<0.01	
POST - PRE															
Self	-0.31*	-0.40*	0.09	-0.24*	-0.38*	0.14	-0.21*	-0.59*	0.38*	-0.26*	-0.46*	0.20*	0.61	0.34	
Others	-0.05	-0.01	-0.04	-0.07	-0.10	0.03	-0.23*	-0.39*	0.16	-0.11*	-0.17*	0.06	<0.01	<0.01	

Physics 2:

	Semester J2-2			Semester N2-2			Semester G2-4			All Semesters			ANOVA <i>p</i> value		
	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F
	PRE														
Self	3.50	2.63	0.87*	3.72	3.18	0.54*	3.70	3.10	0.60*	3.64	2.93	0.71*	0.03	<0.01	
Others	3.32	2.84	0.48*	3.39	2.95	0.44*	3.56	3.10	0.46*	3.43	2.95	0.48*	<0.01	0.27	
POST															
Self	3.63	2.45	1.18*	3.63	2.95	0.68*	3.69	2.96	0.73*	3.65	2.75	0.90*	0.76	<0.01	
Others	3.41	2.73	0.68*	3.42	3.11	0.31*	3.61	3.09	0.52*	3.48	2.95	0.53*	0.02	0.02	
POST - PRE															
Self	0.13*	-0.18*	0.31*	-0.09	-0.23*	0.14	-0.01	-0.14	0.13	0.01	-0.18*	0.19*	0.01	0.85	
Others	0.09*	-0.11	0.20*	0.03	0.16	-0.13	0.05	-0.01	0.06	0.05*	0.00	0.05	0.59	0.16	

Identity as a function of student declared major:

For this analysis, I had to combine all semesters of each course. Still, the number of female physics majors is small. There were 16 female physics majors in Physics 1 and 10 female physics majors in Physics 2. For males and females in each course, and for each variable, we used an ANOVA analysis to test if the average scores were different by major with gender. The ANOVA *p* values are given at the bottom of each list of majors.

	Self-Identity						Recognition by Others					
	Physics 1			Physics 2			Physics 1			Physics 2		
	M	F	M-F	M	F	M-F	M	F	M-F	M	F	M-F
PRE												
Physics	4.19	4.13	0.06	4.40	4.60	-0.20	3.81	3.67	0.14	3.97	4.10	-0.13
Engineering	3.70	3.29	0.41*	3.63	2.97	0.66*	3.50	3.25	0.25*	3.51	3.12	0.39*
Other Science	3.45	2.78	0.67*	3.26	2.69	0.57*	2.97	2.69	0.28*	2.97	2.56	0.41*
Non-Science	3.65	3.17	0.48	3.75	2.50	1.25*	3.28	2.83	0.45	3.06	2.63	0.43
Undeclared	3.59	3.16	0.43*	3.71	3.50	0.21	3.26	2.97	0.29	3.33	3.33	0.00
ANOVA <i>p</i> value	<0.01	<0.01		<0.01	<0.01		<0.01	<0.01		<0.01	<0.01	
POST												
Physics	4.05	3.66	0.39	4.35	4.25	0.10	3.76	3.63	0.13	3.96	3.95	0.01
Engineering	3.53	2.92	0.61*	3.66	2.82	0.84*	3.40	3.07	0.33*	3.55	3.10	0.45*
Other Science	2.92	2.16	0.76*	3.27	2.47	0.80*	2.80	2.45	0.35*	3.11	2.62	0.49*
Non-Science	3.19	2.54	0.65*	3.71	2.41	1.30*	2.99	2.71	0.28	3.14	2.66	0.48
Undeclared	3.26	2.90	0.36	3.63	3.17	0.46	3.16	2.99	0.17	3.43	3.03	0.40
ANOVA <i>p</i> value	<0.01	<0.01		<0.01	<0.01		<0.01	<0.01		<0.01	<0.01	
POST – PRE												
Physics	-0.14	-0.47	0.33	-0.05	-0.35	0.30	-0.05	-0.04	-0.01	-0.01	-0.15	0.14
Engineering	-0.17*	-0.37	0.20*	0.03	-0.15	0.18*	-0.10*	-0.18*	0.08	0.04	-0.02	0.06
Other Science	-0.53*	-0.62*	0.09	0.01	-0.22	0.23	-0.17*	-0.24*	0.07	0.14	0.06	0.08
Non-Science	-0.46*	-0.63*	0.17	-0.04	-0.09	0.05	-0.29*	-0.12	-0.27	0.08	0.03	0.05
Undeclared	-0.33*	-0.26	-0.07	-0.08	-0.33*	0.25	-0.10	0.02	-0.12	0.10	-0.30	0.40
ANOVA <i>p</i> value	<0.01	0.34		0.83	0.93		0.31	0.51		0.68	0.78	

Appendix B: Regression Models for Exams in Physics 1 and Physics 2

Coefficient estimates and multiple regression model statistics for each regression model, for students in Physics 1. The dependent variable is the average exam score. *B* is the un-standardized coefficient, β is the standardized coefficient.

	Model 1			Model 2			Model 3			Model 4		
Model-Level Statistics												
Multiple R^2	0.020			0.565			0.569			0.575		
<i>F</i> statistic <i>p</i> value	<0.001			<0.001			<0.001			<0.001		
Std. Error	14.1			9.4			9.4			9.3		
Independent Variables												
	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.
Constant	69.8		<0.01	64.6		<0.01	61.8		<0.01	56.3		<0.01
<i>Female</i>	-4.44	-0.14	<0.01	-0.11	0	0.89	0.24	0.01	0.76	0.54	0.02	0.50
FMCE Pretest				0.25	0.47	<0.01	0.25	0.45	<0.01	0.24	0.44	<0.01
Math Score				5.43	0.36	<0.01	5.38	0.36	<0.01	5.12	0.34	<0.01
Semester J1-2				-4.69	-0.16	<0.01	-4.51	-0.16	<0.01	-4.60	-0.16	<0.01
Semester K1-2				-11.30	-0.35	<0.01	-10.89	-0.34	<0.01	-10.65	-0.33	<0.01
Pre-Self-Identity							1.34	0.09	<0.01	0.27	0.02	0.53
Pre-Recognition by Others							-0.62	-0.04	0.24			
Pre-Emotional Response										1.94	0.10	<0.01

Coefficient estimates and multiple regression model statistics for each regression model, for students in Physics 2. The dependent variable is the average exam score. *B* is the un-standardized coefficient, β is the standardized coefficient.

	Model 1			Model 2			Model 3			Model 4		
Model-Level Statistics												
Multiple R^2	0.050			0.218			0.250			0.285		
<i>F</i> statistic <i>p</i> value	<0.001			<0.001			<0.001			<0.001		
Std. Error	13.1			11.9			11.7			11.4		
Independent Variables												
	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.	<i>B</i>	β	Sig.
Constant	73.3		<0.01	71.5		<0.01	62.5		<0.01	54.6		<0.01
<i>Female</i>	-6.81	-0.22	<0.01	-5.49	-0.18	<0.01	-3.97	-0.13	<0.01	-3.27	-0.11	<0.01
Math Score				5.42	0.39	<0.01	4.80	0.35	<0.01	4.28	0.31	<0.01
Semester N2-2				4.24	0.15	<0.01	3.67	0.13	<0.01	3.68	0.13	<0.01
Semester G2-4				0.86	0.03	0.36	0.27	0.01	0.77	1.11	0.04	0.23
Pre-Self-Identity							1.56	0.13	<0.01	0.89	0.07	0.04
Pre-Recognition by Others							1.08	0.08	0.06			
Pre-Performance SE										3.55	0.22	<0.01

CHAPTER 6 APPENDICES

Appendix A: Affirmation and Control Exercises and TA Script

Name: _____

Date: _____

Read this list of personal values and think about each of the values. Then circle the two or three values that are MOST important to you. We understand that many of these values may be important to you. Even if you feel that many of the values are important, please pick **only TWO or THREE** of them to circle. There are no right or wrong answers.

The **most** important values to me are: (circle two or three)

Sports Ability

Sense of Humor

Spiritual or Religious Values

Relationships with Friends or Family

Government or Politics

Learning and Gaining Knowledge

Music

Belonging to a Social Group

(such as your community, racial group, or school club)

Independence

Career

Being Smart or Getting Good Grades

Creativity

Being Good at Art

(Please turn the page)

Again, look at the values you picked as most important. List the top two reasons why these values are important to you:

1.

2.

Make a check mark to show how much you agree with each of the following statements:

1. In general, I try to live up to these values.

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
_____	_____	_____	_____	_____	_____

2. These values are an important part of who I am.

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
_____	_____	_____	_____	_____	_____

3. I care about these values.

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
_____	_____	_____	_____	_____	_____

Name: _____

Date: _____

Read this list of personal values and think about each of the values. Then circle the two or three values that are LEAST important to you. Even if you feel that several of the values are not very important, please pick **only TWO or THREE** of them to circle. There are no right or wrong answers.

The least important values to me are: (circle two or three)

Sports Ability

Sense of Humor

Spiritual or Religious Values

Relationships with Friends or Family

Government or Politics

Learning and Gaining Knowledge

Music

Belonging to a Social Group

(such as your community, racial group, or school club)

Independence

Career

Being Smart or Getting Good Grades

Creativity

Being Good at Art

(Please turn the page)

Again, look at the values you picked as least important. List the top two reasons why **someone else** (like another student at your school or a person you've heard about) would pick these as their most important value:

1.

2.

Make a check mark to show how much you agree with each of the following statements:

1. In general, some people try to live up to these values.

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
_____	_____	_____	_____	_____	_____

2. These values are important to some people.

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
_____	_____	_____	_____	_____	_____

3. Some people care about these values.

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
_____	_____	_____	_____	_____	_____

Instructions for TAs.

Thank you very much for your help. Below are the instructions for your recitation. Please read all these instructions before your section so that you can be familiar with them. If you have any questions, you can email or call Lauren Kost-Smith at Lauren.Kost@colorado.edu, (303) 735-0458, Duane F1025.

1. Please read the script OUT LOUD to yourself or someone else 2 or more times sometime before Thurs so that you can read the material naturally in class. It is important not to go off script unless necessary.
2. When recitation is done, keep these instructions out of view so the students from the prior or the next class do not see them.
3. As much as possible, please read WORD FOR WORD all text that is highlighted/bolded in the script.
4. Emphasize underlined text, for example, by reading more slowly and adding emphasis as you speak.
5. Text that is in brackets provides directions to you and should not be read out loud.
6. Please use a clock or watch to make sure that all students are given 15 minutes to complete the writing exercise.
7. If a student arrives after you have finished the instructions, make their arrival time on the outside of the envelope when you give it to them. You can say something like “The rest of the students are working on something. All the directions are inside.”
 - a. If a student comes very late, and would 5 minutes or less to work on the writing exercise, you can just give them the alternative writing packet or ask them just to wait because you will be finishing with this in just a few minutes.
8. It’s important that people be quiet as they complete the writing exercise and approach you with questions, instead of asking them aloud, since talking aloud could call attention to differences in the exercises (which could undermine the study’s validity). If students raise their hands with

questions, wave them up front to ask you the question personally.

9. When students complete the writing exercise, have them put it back in their envelope and then move on to the test.
10. Collect all students' envelopes at the end of the 15 min period and leave them up front with the rest of the class materials. PLEASE DO NOT open or read the contents of any envelopes, to ensure student privacy. An observer who is part of the research team will retrieve the envelopes from you immediately after class, once the last student has left.
11. As students leave, give them a blank copy of the consent form for them to take with them. As you are introducing the FMCE, you might mention that there is a handout for them to take with them on their way out.
12. Materials you will need:
 - a. 1 manila envelope for each student. These will be left for you in your classroom.
 - b. Small stack of alternative assignments. This will be for students who are under 18 or do not want to do the writing exercise. You do not need to hand these out. Students who need them will come up to ask you for one. These will be provided to you by Professor _____.

CLASSROOM SCRIPT [Do not read bracketed sections]

I am going to be reading some of the instructions for today's class to make sure that I say everything I need to.

Today in recitation you'll be doing two different things: a writing exercise and a conceptual survey.

First you'll do the writing exercise.

As Professor _____ mentioned in lecture, one of the things that is important in physics and physics training is clear thinking.

Professor _____ has teamed with some other researchers on campus to develop a writing exercise that will ask you to think about values that are important to people. As you know, thinking and writing clearly is valuable in science, and this exercise is intended to start the process in this course.

I am now going to pass out materials for the writing exercise. All the instructions for the writing exercise are inside the envelope. Please be QUIET as you do this. If you have any questions, please come see me up front. [pass out envelopes]

[LOOK AT CLOCK OR WATCH. BE SURE TO GIVE STUDENTS ~ 15 MINUTES TO WORK ON EXERCISE.]

AFTER THEY ARE DONE ~ 15 MINUTES

Okay, please put your exercise back in the envelope and pass them in. It's OK if you did not complete the whole writing exercise. We'll now move on to the conceptual survey.

[Collect envelopes and any alternative assignment packets. A member of the research team will take them once class is over. Please do not open envelopes to ensure privacy.]

Suggested Responses to Student Questions:

1. Is this for the whole class?—**All recitations will be doing this today.**
2. Do I have to do this?—**Your opinions are important, so I hope you will. Most students will be doing this but the choice is yours.**
3. Why do I have to do this?—**Because clear thinking is important in physics training and this exercise is intended to start that process by encouraging you to think about things that are important to you.**
4. Will I get a grade on this?—**It's not graded.**
5. Why do I have different questions from him/her?—**Some people have different questions from other people. There are a lot of questions to ask, but there's not enough time for everyone to answer all of them.**
6. Can I write on a topic that's not on the list?— **For now, just choose one of the ones on the list.**
7. Why do we get envelopes?—**It's a way of keeping your answers private.**
8. Who will read these?—**Your answers will be private and confidential. They will be shown to the instructor and research team but all names will be removed from the exercise, and replaced with a random ID number, before this occurs.**
9. Why did I have to sign a consent form? **Your answers will be tabulated by other researchers who are working with your professor, and we are required by law to get your consent to allow this.**
10. Why do I need to be 18 for this? **Your answers will be tabulated by other researchers who are working with you professor, and we are required to get your consent to allow this. The law says that to give consent, you have to be 18.**

If students have questions you cannot answer or seem to want more information during the exercise, suggest that they continue with the writing exercise, then take a copy of the consent form with them when they leave. It has contact information for a researcher associated with the project who can answer their questions in more detail.

Appendix B: Attitudes Toward Science Survey

The following items were included on the survey that students took in the second week of the course. We used item #3 to measure students' level of stereotype endorsement.

Please rate the extent to which you agree or disagree with each of the following statements:

- 1) I think about the physics I experience in everyday life.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 2) After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 3) According to my own personal beliefs, I expect men to generally do better in physics than women.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 4) I am not satisfied until I understand why something works the way it does.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 5) I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 6) I study physics to learn knowledge that will be useful in my life outside of school.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 7) I think my physics teachers expect women to do better than men in physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

- 8) If I get stuck on a physics problem on my first try, I usually try to figure out a different way that works.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

9) Nearly everyone is capable of understanding physics if they work at it.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

10) According to general beliefs in society, men are expected to be better at physics than women.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

11) If I don't remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally!) to come up with it.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

12) There are too few women in my physics classes.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

13) If I want to apply a method used for solving one physics problem to another problem, the problems must involve very similar situations.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

14) I enjoy solving physics problems.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

15) I think women typically get better grades than men in physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

16) In physics, mathematical formulas express meaningful relationships among measurable quantities.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

17) Learning physics changes my ideas about how the world works.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

18) I think my physics teachers expect men to do better than women in physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

19) Reasoning skills used to understand physics can be helpful to me in my everyday life.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

20) I can usually figure out a way to solve physics problems.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

21) According to my own personal beliefs, I expect women to do better than men in physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

22) The subject of physics has little relation to what I experience in the real world.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

23) To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

24) I think men typically get better grades than women in physics.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

25) If I get stuck on a physics problem, there is no chance I'll figure it out on my own.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

26) I feel more comfortable talking about physics with students of the same sex than talking about physics with students of the opposite sex.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

27) When studying physics, I relate the important information to what I already know rather than just memorizing the way it is presented.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
-------------------	-------------------	---------	----------------	----------------

The following questions have been designed to investigate ideas about intelligence. There are no right or wrong answers. We are interested in your ideas.

Using the scale below, please indicate the extent to which you agree or disagree with each of the following statements:

28) You have a certain amount of physics intelligence, and you really can't do much to change it.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

29) Your physics intelligence is something about you that you can't change very much.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

30) To be honest, you can't really change how intelligent you are in physics.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

31) You can learn new things, but you can't really change your basic physics intelligence.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

32) People in my physics class believe that people have a certain amount of physics intelligence and they can't really do much to change it.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

33) People in my physics class believe that people's physics intelligence is something that you can't change very much.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

34) People in my physics class believe that you can't really change how intelligent you are in physics.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

35) People in my physics class believe that you can learn new things, but you can't really change your basic physics intelligence.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

Please rate the extent to which you agree or disagree with each of the following statements:

36) I feel like I belong in physics.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

37) People in physics accept me.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

38) I feel like an outsider in physics.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

39) People in physics are a lot like me.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

40) It is important to me to perform well on physics tests.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

41) Having strong physics skills is important to me.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

42) Doing well on physics tests is important to my self-esteem.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

43) My gender is an important part of who I am.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

44) I feel the most comfortable with people from my own gender group.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

45) I have a strong sense of belonging to my gender group.

Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
-------------------	----------	-------------------	----------------	-------	----------------

Appendix C: Table of Raw and Adjusted Average Values of Outcome Measures

This table contains the raw (covariate unadjusted) and covariate adjusted means of each outcome measure used in the initial study. All scores except for the FMCE scores are adjusted based on students' prior math score. FMCE pre-test scores are not adjusted and FMCE post-test scores are adjusted based on FMCE pre-test scores. Standard errors are given in parentheses.

Outcome Measure	Raw Means		Covariate Adjusted Means	
	Males	Females	Males	Females
Overall Mean Exam Score (%)				
Affirmation	69.4 (1.0)	65.2 (1.7)	69.2 (0.9)	67.0 (1.6)
Control	72.7 (1.2)	62.7 (1.7)	72.3 (1.1)	61.3 (1.9)
Exam 1 Score (%)				
Affirmation	76.6 (1.1)	72.6 (1.8)	76.4 (1.0)	74.0 (1.9)
Control	79.0 (1.4)	70.8 (2.2)	78.4 (1.3)	68.5 (2.2)
Exam 2 Score (%)				
Affirmation	69.0 (1.1)	67.1 (1.9)	68.9 (1.0)	70.1 (1.9)
Control	72.0 (1.3)	63.9 (2.0)	71.6 (1.3)	61.9 (2.2)
Exam 3 Score (%)				
Affirmation	61.8 (1.3)	54.3 (2.0)	61.3 (1.2)	55.6 (2.2)
Control	66.5 (1.7)	54.7 (2.4)	66.1 (1.5)	54.5 (2.6)
Final Exam Score (%)				
Affirmation	70.4 (1.1)	66.7 (1.9)	70.4 (1.0)	68.5 (1.8)
Control	73.3 (1.2)	61.3 (2.0)	73.2 (1.3)	60.2 (2.1)
Final Course Score (%)				
Affirmation	73.9 (0.8)	70.5 (1.5)	73.7 (0.7)	72.3 (1.4)
Control	76.0 (1.0)	69.3 (1.4)	75.7 (0.9)	68.2 (1.6)
FMCE Pre-test Score (%)				
Affirmation	39.1 (2.2)	25.9 (3.4)		
Control	41.5 (2.9)	23.7 (4.0)		
FMCE Post-test Score (%)				
Affirmation	72.7 (2.3)	63.6 (4.1)	71.5 (2.0)	74.7 (3.8)
Control	74.7 (3.2)	56.2 (3.9)	72.4 (2.7)	62.0 (4.2)
CAPA Homework Score (%)				
Affirmation	91.8 (0.8)	88.9 (1.5)	91.7 (0.9)	89.3 (1.4)
Control	89.7 (1.4)	94.0 (1.0)	89.7 (1.2)	94.2 (1.8)
Tutorial Homework Score (%)				
Affirmation	74.7 (1.2)	76.7 (1.9)	74.5 (1.1)	77.3 (1.8)
Control	75.4 (1.6)	81.3 (1.6)	75.4 (1.5)	81.6 (2.3)
Participation Score (%)				
Affirmation	81.1 (1.1)	82.5 (1.8)	81.1 (1.1)	83.3 (1.8)
Control	79.6 (1.7)	85.7 (1.6)	79.9 (1.5)	86.0 (2.3)
Clicker Score (%)				
Affirmation	85.2 (1.7)	82.9 (2.6)	85.1 (1.7)	83.5 (2.8)
Control	78.5 (2.6)	88.6 (2.2)	78.9 (2.2)	88.8 (3.4)

This table contains the raw (covariate unadjusted) and covariate adjusted means of each outcome measure used in the replication study. All scores except for the FMCE scores are adjusted based on students' prior math score. FMCE pre-test scores are not adjusted and FMCE post-test scores are adjusted based on FMCE pre-test scores. Standard errors are given in parentheses.

Outcome Measure	Raw Means		Covariate Adjusted Means	
	Males	Females	Males	Females
Overall Mean Exam Score (%)				
Affirmation	63.3 (1.3)	55.3 (1.9)	62.8 (1.1)	56.8 (1.7)
Control	63.3 (1.6)	61.5 (3.3)	63.8 (1.3)	60.1 (3.0)
FMCE Post-test Score (%)				
Affirmation	64.1 (3.0)	49.8 (4.2)	61.2 (2.5)	65.8 (4.5)
Control	68.8 (3.8)	58.6 (7.0)	65.6 (3.1)	66.3 (6.0)

Appendix D: Regression Models for Each Exam

Multiple regression coefficient estimates and significance levels. The dependent variable is the first exam score. The independent variables significantly predicted students' exam 1 scores [$F(1,387)=10.70$, $p<0.01$]. This model accounts for 23% of the variance in mean overall exam score ($R^2=0.23$).

Variables	β	$t(387)$	Significance
Gender	-0.19	-3.65	<0.01
Condition	0.06	1.02	0.31
Stereotype Endorsement	-0.14	-2.37	0.02
Math Score	0.37	7.33	<0.01
Gender \times Condition	0.12	2.21	0.03
Gender \times Stereotype Endorsement	-0.12	-2.05	0.04
Gender \times Math Score	-0.02	-0.36	0.72
Condition \times Stereotype Endorsement	0.12	2.05	0.04
Condition \times Math Score	0.08	1.65	0.10
Stereotype Endorsement \times Math Score	-0.08	-1.61	0.11
Gender \times Condition \times Stereotype Endorsement	0.18	3.04	<0.01

Multiple regression coefficient estimates and significance levels. The dependent variable is the second exam score. The independent variables significantly predicted students' exam 2 scores [$F(1,387)=10.44$, $p<0.01$]. This model accounts for 23% of the variance in mean overall exam score ($R^2=0.23$).

Variables	β	$t(387)$	Significance
Gender	-0.13	-2.53	0.01
Condition	0.09	1.66	0.01
Stereotype Endorsement	-0.09	-1.49	0.14
Math Score	0.36	7.00	<0.01
Gender \times Condition	0.18	3.26	<0.01
Gender \times Stereotype Endorsement	-0.03	-0.58	0.56
Gender \times Math Score	-0.05	-0.90	0.37
Condition \times Stereotype Endorsement	0.18	2.96	<0.01
Condition \times Math Score	0.06	1.28	0.20
Stereotype Endorsement \times Math Score	-0.11	-2.31	0.02
Gender \times Condition \times Stereotype Endorsement	0.21	3.54	<0.01

Multiple regression coefficient estimates and significance levels. The dependent variable is the third exam score. The independent variables significantly predicted students' exam 3 scores [$F(1,387)=11.79, p<0.01$]. This model accounts for 25% of the variance in mean overall exam score ($R^2=0.25$).

Variables	β	$t(387)$	Significance
Gender	-0.22	-4.36	<0.01
Condition	-0.05	-0.94	0.35
Stereotype Endorsement	-0.09	-1.57	0.12
Math Score	0.39	7.80	<0.01
Gender \times Condition	0.08	1.48	0.14
Gender \times Stereotype Endorsement	-0.04	-0.70	0.48
Gender \times Math Score	-0.04	-0.80	0.42
Condition \times Stereotype Endorsement	0.10	1.71	0.09
Condition \times Math Score	0.02	0.51	0.61
Stereotype Endorsement \times Math Score	-0.10	-2.04	0.04
Gender \times Condition \times Stereotype Endorsement	0.04	0.64	0.52

Multiple regression coefficient estimates and significance levels. The dependent variable is the final exam score. The independent variables significantly predicted students' final exam scores [$F(1,387)=12.28, p<0.01$]. This model accounts for 26% of the variance in mean overall exam score ($R^2=0.26$).

Variables	β	$t(387)$	Significance
Gender	-0.23	-4.62	<0.01
Condition	0.09	1.70	0.09
Stereotype Endorsement	-0.13	-2.26	0.02
Math Score	0.37	7.51	<0.01
Gender \times Condition	0.19	3.45	<0.01
Gender \times Stereotype Endorsement	-0.02	-0.25	0.80
Gender \times Math Score	-0.01	-0.26	0.80
Condition \times Stereotype Endorsement	0.12	2.07	0.04
Condition \times Math Score	0.11	2.38	0.02
Stereotype Endorsement \times Math Score	-0.04	-0.83	0.41
Gender \times Condition \times Stereotype Endorsement	0.13	2.12	0.04

Appendix E: Regression Models for Homework, Participation and Clicker Scores

The dependent variable is the CAPA homework score. The independent variables significantly predicted students' CAPA homework scores [$F(1,392)=2.66, p=0.02$] and account for 4% of the variance ($R^2=0.04$).

Variables	β	$t(392)$	Significance
Gender	0.04	0.78	0.44
Condition	-0.06	-1.09	0.28
Math Score	0.12	2.09	0.04
Gender \times Condition	-0.15	-2.60	0.01
Gender \times Math Score	0.03	0.60	0.55
Condition \times Math Score	0.08	1.48	0.14

The dependent variable is the *Tutorial* homework score. The independent variables significantly predicted students' *Tutorial* homework scores [$F(1,392)=3.58, p<0.01$] and account for 5% of the variance ($R^2=0.05$).

Variables	β	$t(392)$	Significance
Gender	0.13	2.59	0.01
Condition	-0.08	-1.52	0.13
Math Score	0.15	2.68	0.01
Gender \times Condition	-0.06	-1.00	0.32
Gender \times Math Score	0.02	0.37	0.72
Condition \times Math Score	0.09	1.80	0.07

The dependent variable is the participation score. The independent variables significantly predicted students' participation scores [$F(1,392)=3.19, p<0.01$] and account for 5% of the variance ($R^2=0.05$).

Variables	β	$t(392)$	Significance
Gender	0.12	2.39	0.02
Condition	-0.03	-0.48	0.63
Math Score	0.08	1.44	0.15
Gender \times Condition	-0.06	-1.13	0.26
Gender \times Math Score	0.12	2.28	0.02
Condition \times Math Score	0.13	2.55	0.01

The dependent variable is the clicker score. The independent variables significantly predicted students' clicker scores [$F(1,392)=2.25, p=0.04$] and account for 3% of the variance ($R^2=0.03$).

Variables	β	$t(392)$	Significance
Gender	0.08	1.58	0.12
Condition	0.01	0.19	0.85
Math Score	0.03	0.48	0.63
Gender \times Condition	-0.12	-2.19	0.03
Gender \times Math Score	0.08	1.47	0.14
Condition \times Math Score	0.09	1.69	0.09