THE EFFECTIVENESS OF INTERACTIVE ENGAGEMENT IN INTRODUCTORY PHYSICS COURSES AT THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

David Tyler Guynn

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Physics and Astronomy.

Chapel Hill
2016

Approved by:

Laurie E. McNeil

Alice D. Churukian

Daniel E. Reichart

ABSTRACT

DAVID TYLER GUYNN: THE EFFECTIVENESS OF INTERACTIVE ENGAGEMENT IN INTRODUCTORY PHYSICS COURSES AT THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL.
(Under the direction of Laurie E. McNeil)

The Department of Physics and Astronomy at the University of North Carolina at Chapel Hill (UNC-CH) has implemented interactive-engagement in its introductory physics curriculum for students majoring in both the life sciences and the physical sciences. As a measure of teaching effectiveness, UNC-CH has been administering both the Force Concept Inventory (FCI) and Conceptual Survey of Electricity and Magnetism (CSEM) to students in introductory courses since Fall 2007.

This project examines students' performance on both the FCI ($N = 7863$) and CSEM ($N = 5222$) using several established metrics for determining learning gains. This study finds that the implementation of interactive-engagement has a statistically significant increase of learning gains, independent of student gender or ethnicity. Furthermore, it was observed that learning gaps between genders narrowed for students enrolled in courses for physical science majors. Finally, further opportunities for study and data analysis are described for future use within UNC-CH's research program.

To Adaline.

ACKNOWLEDGMENTS

This thesis would not have been possible without the direct or indirect contributions of the following individuals and groups:

My wife Natalie: for your unwavering love and support over the past seven years, and for being a wonderful mother to our child(ren).

Alice: for your trust, advice, and support throughout my entire PAER project (and time at UNC). I couldn't have done it without you.

Laurie: for driving the implementation of interactive-engagement and the use of PAER findings at UNC.

The PAER group at UNC: I greatly appreciate and acknowledge your hard work in implementing interactive-engagement education at UNC and collecting the data used in this project.

All the faculty and graduate students who teach introductory physics every day at UNC: thank you for your contributions to undergraduate physics education and for making these learning gains possible.

To my father and grandparents: for teaching me the value of an education.

Matt, Rama, and Rich: for seeing my potential.

Roger: For helping me get started on the journey – forever appreciative of our conversations and friendship from the fire department until today.

To all my friends and comrades in the fire service: I wouldn't be who I am today without you all, especially Lt. Mike Jenkins, Chief Bobbie Slayton, Capt. Chris Trussler, Capt. Lynn Flora, Capt. Clarence Turpin, FF Jacob Palmer, Lt. Andy Foley, Mickey Upham, FF/PM David Palmer and Lt. Aaron Murray. To Captain Phil Dillon: Phil, thanks for helping me learn how to finish what I start.

TABLE OF CONTENTS

1 INTRODUCTION 1
1.1 Efforts at the University of North Carolina at Chapel Hill 2
1.2 Personal Motivation 4
1.3 Research Questions 4
2 LITERATURE REVIEW 6
2.1 Blank slates and misconceptions 6
2.2 Piaget and Constructivism 10
2.2.1 Social Constructivism 13
2.2.2 Where are our students in Piaget's model? 14
2.3 Teaching models and interactive pedagogies 16
2.3.1 SCALE-UP 18
2.3.2 New Studio 20
3 METHODOLOGY 22
3.1 Written Concept Surveys 22
3.1.1 Force Concept Inventory 22
3.1.2 Conceptual Survey of Electricity and Magnetism 23
3.1.3 Diagnostic Test Administration and Recordkeeping 24
3.2 University Registrar Demographic Data 24
3.3 Statistical Analysis 26
3.3.1 Regression Analysis 26
3.3.2 Analysis of Variance 28

3.3.3 Post hoc testing 29
3.4 Statistical Software 30
4 DATA AND ANALYSIS 31
4.1 Measuring learning gains 31
4.1.1 Normalized gain 31
4.1.2 Normalized change 32
4.2 Elimination of Outliers 33
4.3 Force Concept Inventory Analysis 36
4.3.1 Outlier and Metric Analysis 36
4.3.2 Demographics 36
4.3.3 Learning gains 38
4.3.4 FCI learning gains and gender 42
4.3.5 FCI learning gains and ethnicity 45
4.4 Conceptual Survey of Electricity and Magnetism Analysis 48
4.4.1 Outlier and Metric Analysis 48
4.4.2 Demographics 49
4.4.3 Learning gains 50
4.4.4 CSEM learning gains and gender 52
4.4.5 CSEM learning gains and ethnicity 55
5 RESULTS AND CONCLUSIONS 60
5.1 Has the implementation of interactive-engagement pedagogy at UNC-CH led to student gains in understanding? 60
5.2 Has the implementation of interactive-engagement pedagogy at UNC-CH led to a significant change in understanding with regard to race or gender? 62
5.3 How do the results at UNC-CH compare to other institutions? 63
5.4 Future Work and Conclusion 64

A INTRODUCTORY PHYSICS COURSE TOPICS 65

B UNC-CHAPEL HILL PEER INSTITUTIONS 70

LIST OF TABLES

2.1 Summary of Piaget's stages of cognitive development 11

3.1 Description of database fields from registrar 25

4.1 Possible normalized gains by pre-test score 35
4.2 Ethnicities in FCI datasets 38
4.3 Ethnicities in FCI datasets with combined ethnic factors 38
4.4 FCI-measured learning gains by gender and modality, IPPS 42
4.5 FCI-measured earning gains by gender and modality, IPLS 43
4.6 FCI-measured learning gains by ethnicity and modality, IPPS 45
4.7 FCI-measured learning gains by ethnicity and modality, IPLS 46
4.8 Ethnicities in CSEM datasets 50
4.9 Ethnicities in CSEM datasets with combined ethnic factors 50
4.10 CSEM-measured learning gains by gender and modality, IPPS 53
4.11 CSEM-measured learning gains by gender and modality, IPLS 54
4.12 CSEM-measured learning gains by ethnicity and modality, IPPS 56
4.13 CSEM-measured learning gains by ethnicity and modality, IPLS 57

A.1 Course topics: IPLS Mechanics (PHYS 104/114)............................. 66
A.2 Course topics: IPPS Mechanics (PHYS 116/118)............................. 67
A.3 Course topics: IPLS Electricity and Magnetism (PHYS 105/115) 68
A.4 Course topics: IPPS Electricity and Magnetism (PHYS 116/118) 69

LIST OF FIGURES

4.1 Regression results of g versus c, with and without identified outliers 37
4.2 Enrollment in mechanics courses by gender 39
4.3 FCI-measured learning gains by term, IPPS courses 40
4.4 FCI-measured learning gains by term, IPLS courses 41
4.5 FCI-measured learning gains by gender and modality, IPPS 43
4.6 FCI-measured learning gains by gender and modality, IPLS 44
4.7 FCI-measured learning gains by ethnicity and modality, IPPS 46
4.8 FCI-measured learning gains by ethnicity and modality, IPLS 47
4.9 Regression results of g versus c, with and without identified outliers 49
4.10 Enrollment in E&M courses by gender 51
4.11 CSEM-measured learning gains by term, IPPS 52
4.12 CSEM-measured learning gains by term, IPLS 53
4.13 CSEM-measured learning gains by gender and modality, IPPS 54
4.14 CSEM-measured learning gains by gender and modality, IPLS 55
4.15 CSEM-measured learning gains by ethnicity and modality, IPPS 56
4.16 CSEM-measured learning gains by ethnicity and modality, IPLS 57

1 INTRODUCTION

If you watch the evening news on a regular basis, you will see almost weekly stories discussing the failure of the American educational system to produce students fluent in the fields of science, technology, engineering and mathematics (STEM). Most of these stories are based on either the relative decline in performance of American high school students on standardized testing, or the paucity of students completing majors within the STEM fields at American universities. Although much of the blame for these failures has been placed at the high school level, it is acknowledged that the quality of undergraduate science education plays a large role in the development of secondary school science teachers (McCormick 2004). Furthermore, introductory science courses for undergraduates are considered to be where students learn to "love and appreciate" or hate science (McCormick 2004).

What, then, do students think about introductory science education? In one 1994 survey, students disliked the fact that the majority of science courses consisted mostly of passive lectures, and that classes were tedious and mainly required memorization of large amounts of information (Seymour 1995). In another study, students found introductory physics courses in particular to be "boring, crammed with too much material, narrowly focused on numerical manipulation and computation, and biased against any attention to the 'big picture' of what physics is about" (Knight 2004).

The most important question, however, is whether students actually understand the information being taught in undergraduate science courses. Assessment of student understanding and the improvement of physics teaching has been the driving force behind the development of Physics Education Research (PER). In PER, physicists study physics education from a scientific approach, using the results of systematic studies to determine the effectiveness of physics education (McDermott and Redish 1999). These studies have found that many students leave introductory physics courses with very little understanding of the basic concepts of physics, and that the standard lecture model for teaching physics does very little to improve student understanding. The application of PER to science education has motivated the development of courses that move away from the standard passive lecture format to an interactive-engagement approach – an approach with quantitative data demonstrating its success (Knight 2004).

1.1 Efforts at the University of North Carolina at Chapel Hill

The University of North Carolina at Chapel Hill (UNC-CH), like most institutions, offers two levels of introductory physics, usually known as algebra-based and calculus-based. The algebra-based course sequence usually enrolls students majoring in the life sciences, and many of these students are also on a pre-professional health sciences track. Students who enroll in the calculus-based course sequence generally are majoring in physics, another physical science, or computer science. Historically, all of these courses were taught using the standard lecture/laboratory model, and the calculus-based sequence also included a recitation section usually led by a graduate student.

However, in 2005, UNC-CH received a grant from the National Science Foundation to "transform introductory physics teaching" by implementing new courses based on the findings of PER (McNeil 2004). This project, led by then-chair Dr. Laurie McNeil, examined a variety of pedagogical methods with a goal of a "sustainable transformation" that would allow for the modified courses to continue, without regard to the faculty members or instructors involved. The project was primarily driven by low apparent gains in student understanding in the calculus-based courses, as well as a perceived lack of connection between the material in the lab and lecture sections.

More progress was made with the award of a PhysTEC grant in spring of 2007, with the primary goal of increasing the number of highly-qualified high school physics teachers by implementing a teacher preparation program. The department also made a commitment to implement an instructional approach that was more interactive-engagement in nature. At that point, Dr. Alice Churukian joined the department, and began introducing cooperative group problem solving during the recitation sections as a way to introduce interactive-engagement (Churukian 2016b).

Between 2008 and 2010, discussions centered on adopting either SCALE-UP or the Lecture/Studio format for courses in the calculus-based sequence. Work was ongoing to secure funding for the conversion of classrooms into a setup compatible with interactive-engagement. These renovations were completed in Summer of 2010, and SCALE-UP was implemented as a pilot process in Fall 2010. As a pilot, the department conducted one section of SCALE-UP and the remainder in a traditional format, with the exception of Fall 2011 where a single section of lecture/studio was used. After several semesters, it was apparent that students in the interactive-engagement section performed better on common tests, so it was decided that all sections of introductory courses would convert to interactive-engagement. The department determined that a suitable infrastructure did not exist for a full implementation of SCALE-UP, so Lecture/Studio was implemented for all sections of the calculus-based sequence beginning in Fall 2013.

The original introductory physics sequences (PHYS 104/105 and PHYS 116/117) followed the traditional algebra-based and calculus-based curricula. Beginning in 2014, however, all introductory physics courses were restructured to not only utilize a Lecture/Studio format but also have a new curriculum. Because of the changes in the course material, these courses were renumbered (PHYS 104/105 to 114/115, PHYS 116/117 to 118/119). The changes in the curriculum included the addition of several calculus topics to the new 114/115 course sequence. To eliminate confusion, this thesis will refer to both 104/105 and 114/115 as Introductory Physics for Life Sciences (IPLS). Similarly, this thesis will refer to 116/117 and 118/119 as Introductory Physics for Physical Sciences (IPPS). Details of the course topics for the IPLS and IPPS sequences can be found in Appendix A of this thesis.

1.2 Personal Motivation

My motivation for this project comes from a deeply personal interest in the quality of undergraduate science education. I have a very clear memory from the beginning of my first undergraduate course in electrostatics, when the instructor began a discussion of Gauss' law. She began the discussion by writing the integral form of the law on the board, and then began discussing the electric flux. She drew a sphere on the board, shaded in the surface element $\mathrm{d}\vec{a}$, and then arbitrarily drew electric field vectors throughout the surface of the sphere. I remember becoming incredibly confused at this point, because obviously the flux was important, and she had explained the flux in a way that depended on the number of field vectors per surface area of the sphere. Yet, it appeared from her drawing that the distribution didn't matter at all. Of course, I would learn later that this was simply an abstract construct, and the symmetry of the sphere allowed for the problem to be easily solved. However, there was a clear tension between my concrete understanding and her abstract understanding of electric fields – one that caused a great deal of angst during the formative phases of my physics education[1]. This is precisely the sort of conflict that the application of PER hopes to resolve.

1.3 Research Questions

The scope of the transformation of the introductory physics at UNC-CH has produced a great deal of data. These data should allow the Department of Physics and Astronomy to assess the efficacy of interactive-engagement pedagogy, as well as help other similar institutions justify the transition to active learning in the sciences. This thesis examines three specific questions:

---

[1] I should also note here that I am deeply indebted to this instructor. I never would have completed my undergraduate studies without her encouragement and support, and this is in no way a negative reflection on her style of teaching.

1. Question 1: *Has the implementation of interactive-engagement pedagogy at UNC-CH led to student gains in understanding?*

2. Question 2: *Has the implementation of interactive-engagement pedagogy at UNC-CH led to significant gains in understanding that differ between gender and racial groups?*

3. Question 3: *How do the results at UNC-CH compare to other institutions?*

2 LITERATURE REVIEW

Understanding how students learn is critical to developing a curriculum that engages students and transfers knowledge effectively. The traditional pedagogy of physics education is the lecture, a form of instruction not popular among students (Seymour 1995). Lecturing is considered a transmissionist mode of instruction, where the instructor delivers a lesson and students learn passively. The term lecture comes from the Latin *lectura*, meaning "to read" – historically, books were rare, so a single copy of the book was often read to an entire class. However, a parallel can also be seen from the origin of physics as "natural philosophy": Aristotle's philosophical study of the natural world. The study of natural philosophy in the European universities of the middle ages consisted primarily of "memorizing Aristotle's speculations on the subject, and... having hair-splitting disputations as to their meanings and their possible implications" (Mann 1912). Natural philosophy would be the dominant mode of the study of the natural world until the development of the scientific method in the 19th century (Meltzer and Otero 2015). Even with these changes, the lecture served as the primary tool of physics education.

2.1 Blank slates and misconceptions

Aristotle viewed students as a blank slate – the tabula rasa. Aristotle wrote in *On the Soul* that the mind:

> ...is in a sense potentially whatever is thinkable, though actually it is nothing until it has thought... What it thinks must be in it just as characters may be said to be on a writing-tablet on which as yet nothing stands written: this is exactly what happens with [the] mind

Filling the blank slate is the ostensible goal of the transmissionist mode of instruction. In his 1938 book *Experience and Education*, John Dewey described this traditional mode of education as “bodies of information and of skills that have been worked out in the past... the chief business of the school is to transmit them to the new generation” (1938). If the mind is, in fact, a blank slate, than the transmissionist view of education could be a reasonable way to teach physics.

Unfortunately, students do not usually begin their physics study as a blank slate. Research has shown that students’ prior knowledge determines a majority of what they can learn from a given situation (Fisher 2004). One PER researcher states that:

> Every student begins physics with a well-established system of commonsense beliefs about how the world works derived from years of personal experience... Instruction that does not take [these beliefs] into account is almost totally ineffective (Hestenes et al. 1992).

In short, if new information is presented to a student without the development of context and meaning, and this information conflicts with pre-existing ideas, then the pre-existing ideas will take precedence over instruction.

In the literature, these pre-existing beliefs are known as misconceptions or preconceptions.[1] By definition, a preconception is an idea held before instruction. Some preconceptions are largely in agreement with physical theory and can aid learning, while others – misconceptions – are incorrect and pose a barrier to learning (Clement et al. 1989). Research with students has shown that misconceptions are often shared among many students (Clement 1982). Misconceptions in science education generally have three primary characteristics – they are shared by a significant portion of the population, they produce consistent error patterns, and are resistant to teaching (Fisher 2004). Examples of common misconceptions in physics include, but are certainly not limited to, the concept of force, the

[1] This thesis does not discuss phenomenological primitives (“P-prims”), which also play a role. For more information on P-prims, Chapter 15 of Fischbein (1987) serves as a good introduction.

relationship between force and acceleration, motion implying the existence of a force, non-zero acceleration with zero velocity, and the relationship between work and energy (Arons 1981; Clement 1982; Halloun and Hestenes 1985). All of these misconceptions have been extensively studied through diagnostic testing and student interviews. To provide an example of the relationship between the misconception and the physical concept, I will discuss the "motion implies force" misconception in greater detail.

In their physical world, students know from a young age that they must push on a block in order to keep it moving on a table. This presents an issue when students begin studying Newton's laws of motion. First, students most often consider forces as sensations corresponding with pushing or pulling an object, or of feeling the weight of an object when held in the student's hand, limiting the student's idea of force to one which must be directly perceived in order to be understood (Arons 1981). Because the force of friction is a force which is not directly perceived, students assume that the block comes to a rest on the table because of the absence of a force, rather than the presence of a force. As physics instructors, we want to replace this concept with $F = ma$, but as Hestenes states above, eliminating a misconception requires more than just presenting a formula – it requires presenting Newton's laws of motion with context while considering a student's likely misconceptions.

Various diagnostic tests have been developed in an attempt to understand students' pre-existing misconceptions and determine if they are appropriately addressed by instruction. In 1985, Ibrahim Halloun and David Hestenes of Arizona State University developed two multiple-choice diagnostic tests to probe the impact of students' initial knowledge on learning physics. One test, for mathematics, was used to determine the impact on students' mathematical ability on a physics course. The other, the Mechanics Diagnostic Test, probed students' preconceptions regarding basic mechanics. The mathematics test was only delivered pre-instruction. Topics included algebra, arithmetic, trigonometry, understanding graphs, mathematical reasoning, and calculus. The physics diagnostic test was designed to be delivered both pre- and post-instruction to allow measurement of learning gains in an introductory physics course, and focused on Newtonian mechanics. The concepts covered in the physics diagnostic test had also been previously identified by other researchers as concepts that frequently conflicted with students' pre-existing misconceptions. After administering the test to over 1500 students, Halloun and Hestenes found that (a) the initial knowledge of a student has a large effect on performance in a physics course, and (b) that conventional (lecture-based) instruction produces very little gain in physics knowledge.

Then, in 1992, David Hestenes, along with Malcolm Wells and Gregg Swackhamer, released a paper describing the Force Concept Inventory (FCI) – one of the two diagnostic tests used in this project. The FCI consists of multiple-choice questions that require students to make a choice between Newtonian concepts and common misconceptions. Hestenes states that about 60% of the FCI is the same as the previous Mechanics Diagnostic Test – the primary differences come from alterations that rely on a "more systematic and complete profile of the various misconceptions" (Hestenes et al. 1992). In 1998, Richard Hake of Indiana University published an analysis of the FCI scores for over 6000 introductory physics students. Using a metric to quantitatively assess the ratio of a student's learning gain to the student's possible gain, known as the normalized gain[2], Hake found that, with respect to the normalized gain, students enrolled in a "traditional" physics course performed almost two standard deviations lower than students in an "interactive-engagement" course. In his study, Hake defined an interactive-engagement course to be "designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on and hands-on activities which yield immediate feedback through discussion", and a traditional course that "make(s) little or no use of interactive-engagement methods, relying primarily on passive-student lectures, recipe labs, and algorithmic-problem exams" (Hake 1998). The remainder of this thesis will make the same differentiation between

[2]See Chapter 4, Data, for a more in-depth description of the normalized gain.

traditional and interactive-engagement physics courses.

With the understanding that traditional instruction in introductory physics courses produces very little knowledge gain, and that knowledge gain is hampered by the existence of students' pre-existing misconceptions, the next question is simply to ask "Why?" Arnold Arons makes the statement in a 1981 paper that "without explicit help and guidance, students completely fail to visualize the presence of physical effects which transcend immediate sense perceptions." Another way to say this is that students understand physical objects, but often have problems with a verbal hypothesis independent of a physical object (McKinnon and Renner 1971). While this seems fairly logical at face value, a deeper understanding of the root cause of this finding may be found in the work of Jean Piaget.

2.2 Piaget and Constructivism

Jean Piaget began his study of cognitive development in 1918 while working for the creators of the Binet-Simon intelligence test. While observing children who were taking the test, he found that students of "similar ages made similar types of mistakes, and it occurred to him that... the key to understanding human development is not in what children get wrong, but how they get it wrong" (Jardine 2006). He then noticed that most children, by age twelve, stopped making the same mistakes as the younger children. Piaget was not necessarily interested in the development of children per se, nor was he attempting to answer the philosophical and epistemological question "What is knowledge?" (Jardine 2006). Instead, Piaget wanted to develop an understanding of how knowledge grows, particularly in children. He termed this study genetic epistemology. Specifically, he wanted to develop a genetic epistemology for the growth of knowledge as practiced in the sciences. He believed that the key to understanding the growth of knowledge was to understand how students organized knowledge – their schema.

A schema is an "organized pattern of thought... that organizes categories of information and the relationships among them" (Renner and Lawson 1973). As children grow and

Table 2.1: Summary of Piaget's stages of cognitive development

Stage	Key Characteristics of Learners	Age Range
Sensory-motor	• Object permanence • Focused on the immediate spatial surroundings • Learns through physical activity	0-18 months
Preoperational	• Immersed in language and play • Does not demonstrate understanding of "conservation reasoning" • Child does not reverse thinking • Child centers attention on a particular aspect of an object or situation	18 months- 7 years
Concrete operational	• Can only perform mental operations that are tied to objects (i.e., classifying and ordering) • Can reverse thinking • Beginning of "thought experiments" and inductive reasoning	7 to 11 years
Formal operational	• Can understand mental operations independent of an object • May form and test hypotheses • Decenteration • Capable of logical and mathematical thought	11 to 15+ years

mature, their schemata also grow and mature, directly affecting how children understand new information. Piaget had already found evidence for a model describing development in cognitive reasoning while working with Binet and Simon. Later, in a study of Swiss children, Piaget noticed similarities in children's schema and their ability to learn as a function of age, allowing Piaget to develop a more general model that encompasses four stages of cognitive development, as shown in Table 2.1.

The potential differences between the cognitive levels of students present a challenge: when teaching, it is not enough to make a verbal statement and expect students to have an a priori understanding that allows them to solve a problem. Instead, instructors should understand that students have a variety of schemata and preconceptions that will influence their learning. As Jardine (2006) states:

> [Piaget learned that] children seemed to put things together differently than

> adults. As adults, we have learned these standards so well that we tend to "forget" our long and complex agency in the world that we experience... forgetting that we have, collectively and individually, learned "the way things really are" by collectively and individually making something of our experiences.

The existence of an appropriate schema is also applicable to problem solving in physics. First year physics students often find success in answering questions that have been seen before (von Glasersfeld 1995). Students can easily learn the patterns inherent in "plug-and-chug" questions – test questions requiring only a simple substitution of relevant values. These same students, however, can have a great deal of difficulty in answering a similar problem that has only a slight change, simply because it deviates from the established pattern. These students don't exhibit a true understanding of the problem or the conceptual relationships described by memorized formulae.

Piaget defined two concepts describing how students learn: "assimilation" and "accommodation" (Furth 1969). Assimilation is the process by which students interpret new information through an existing schema, often re-interpreting this information to fit into the pre-existing schema. The possibility also exists for a student to reject new information. If the student's schema is completely incompatible with new information, then they might not be able to understand the new information at all, causing the information to be completely rejected (Jardine 2006).

On the other hand, accommodation is the process of interpreting new information and altering pre-existing schema to be compatible with the new information. Accommodation resolves the discrepancies between the student's current schemata and the new information. Piaget felt that this potential disequilibrium was important, stating that "disequilibrium in cognitive structure is what motivates progress... of cognitive reasoning" (1950). Disequilibrium can provide an additional benefit by creating curiosity in students, further motivating their learning (Johnson et al. 1997).

Piaget's model of learning is not just absorbing new information: instead, it is an active process where the learner must reconstruct their old schema to accommodate new information. In short, knowledge is constructed by the interaction of the assimilation and accommodation processes. Rosalind Driver describes constructivism as:

> Rather than seeing the learner as a passive absorber of information, a constructivist perspective views the learner as actively engaged in constructing meaning, bring his or her prior knowledge to bear on new situations, and, if the purposes are worthwhile, adapting those knowledge structures (1995).

2.2.1 Social Constructivism

Textbooks and lectures often present a picture of science being fixed and absolute, which often serves to place students at an uncomfortable distance from the topic. This view is at odds with a scientist's understanding of what science is: a field of study where theory is provisional, rather than absolute, and requiring both personal and social construction (Driver 1995). Modern science – what some call "big science" – is all about collaboration among large groups of people across the world. A practical example of this is the discovery of the Higgs boson at CERN's Large Hadron Collider, where the paper announcing the results had almost nine pages of collaborators listed (The ATLAS Collaboration 2012). Science education can be provided in a similar fashion in a manner that is compatible with constructivist beliefs by acknowledging that science is "a product of social processes and that learning science requires students to be initiated into this scientific culture" (Driver 1995).

These philosophical ideas about science education can be more formally described within the framework of social constructivism. These ideas were first developed by the Russian psychologist Lev Vygotsky, who believed that higher mental functioning was social in nature (Otero 2004) . In his theory, this is largely in part due to the use of language in learning. Vygotsky drew a difference between spontaneous concepts and scientific concept. Spontaneous concepts consist of intuitive knowledge that the learner gains through experience and has never been explicitly explained through language. Scientific concepts, however, are introduced and developed through language, and require both verbal definitions and conscious comprehension to understand the concept.

Also in the Vygotskian perspective, language has meaning because individuals and society provide meaning, through the efforts of two or more persons. Speaking from the social constructivist perspective, Kenneth Gergen states that " knowledge is achieved when one transcends the particular and grasps the general" (1995). Gergen then continues by making the statement that scientific or mathematical principles can be equated to truth, and truth does not require context. Once these principles are converted to language, however, the learner requires context as they construct their own knowledge. Understanding that language has a social component allows for instructors to consider context in learning. Finally, social constructivist educational theory realizes that there is value in considering the viewpoints and perspectives of other learners (Marzano 2007).

2.2.2 Where are our students in Piaget's model?

Considering the characteristics of both the concrete operational and formal operational stages of Piaget's model, it is reasonable to wonder where introductory physics students lie within the model. Many different diagnostic tasks have been developed to determine a student's cognitive level. While these tasks, often known as Piagetian tasks, are usually administered to students functioning at the preoperational or concrete operational level (Weisman and Safford 1971), they also have been found to have relevance to undergraduate physics education.

In a 1973 paper, John Renner and Anton Lawson presented the results of experiments conducted on three groups of students: randomly selected high school students, first year college students, and law students in their second and third years. They presented these students with two Piagetian tasks. The first task, regarding conservation of volume, determines whether a student is entering the formal operational stage. The second task, asking students to exclude irrelevant variables in an experiment, can determine whether a student is entering the formal operational stage, within the formal operational stage, or capable of using propositional logic.

Renner and Lawson found that, in their population of first year college students ($N = 185$), only 133 (72%) were successful on the conservation of volume task and 77 (42%) were successful on the exclusion task. They also report that, in the high school group ($N = 196$) that only 97 (49%) were successful on the conservation of volume task and 73 (37%) were successful on the exclusion task. Renner and Lawson caution that the test is not an absolute diagnostic of whether a student is formal operational; instead, it allows the inference of a student's ability to utilize propositional logic. Propositional thought, or propositional logic, is the ability to make a logical conclusion based on a written statement rather than a direct observation. Given these findings, the authors assert that a large percentage of the adolescent population – the target population for introductory physics courses – do not function at the formal operational level in Piaget's model.

The relevance of this finding is magnified when considering what scientific reasoning actually is. Lawson suggests that scientific reasoning is entirely hypothetico-deductive in nature, using the scientific method to hypothesize, predict and experiment (Moore 2012). While this seems obvious, hypothetico-deductive reasoning draws upon a particular set of skills: proportional reasoning, probability reasoning, correlation reasoning, and the appropriate use of variables, all of which are very similar in practice to the Piagetian tasks that Renner and Lawson evaluated in their 1972 study (Lawson 1982, 2005). If students are not yet capable of scientific reasoning and propositional thought, it follows that they will have difficulty with a physics course that immediately requires these skills for success.

2.3 Teaching models and interactive pedagogies

There are a variety of teaching models available for use in science education. These models serve as an abstraction between pedagogy, the method and practice of teaching, and cognitive theory, the scientific basis behind knowledge and learning. Models are a useful tool for understanding and evaluating a given pedagogy against the cognitive theories that the model represents.

One such model is the learning cycle (Karplus 1977). The learning cycle is based on Piaget's developmental model and was first described in the early 1960s (Zollman 2004). While this model was originally developed for elementary school students, it has been shown to be effective for older students as well. The learning cycle consists of three phases that "combine experience with social transmission":

1. Exploration
2. Concept introduction
3. Concept application

The exploration phase requires students to "explore a concept by performing a series of activities" while pursuing a general goal (Zollman 2004). This requires students to gain experience through their personal actions and reactions and the actions and reactions of their peers. The learning experience should "raise questions and complexities" that can't be answered with their current schema, causing a disequilibrium in the Piagetian sense.

The concept introduction phase provides a definition of a concept that helps students develop a new schema and resolve the disequilibrium from the exploration phase. It usually consists of an "expository statement of concepts and principles" rather than an experimental activity (Zollman 2004). Concept introduction is most effective when it "involves a formal definition of a concept whose concrete definition is already understood by the students" (Karplus 1977), another application of Piagetian theory. Finally, the concept application phase allows students to apply the new concept (and modified schema) to additional scenarios or situations. This phase is needed to extend applicability of the new concept from concrete to abstract, and provides learners more time for accommodation.

In his book *Teaching Physics with the Physics Suite*, Edward Redish (2003) combines the Piagetian and Vygotskian genetic epistemologies with results from cognitive science research on memory to develop “The Cognitive Model”, which is used to provide a theoretical underpinning for the pedagogies that he describes in his book. Redish summarizes his model into five main principles:

1. The constructivism principle
2. The context principle
3. The change principle
4. The individuality principle
5. The social learning principle

The constructivism principle is equivalent with the previous discussion of constructivism in this work – students build knowledge by making connections to existing knowledge. Redish’s change principle invokes many of Piaget’s ideas regarding assimilation and accommodation, while acknowledging the difficulty inherent in changing students’ pre-existing misconceptions and mental models. As a general principle, Redish claims that “It is reasonably easy to learn something that matches or extends an existing schema, but changing a well-established schema substantially is difficult.”

In the context principle, Redish makes an argument regarding the variety of students’ cognitive responses, stating that “what people construct depends on the context.” Redish’s argument is similar to the argument that Gergen makes regarding language and social constructivism, where the conversion of a physical principle (a “truth”) into language requires that instructors consider context while students process new information (1995).

Redish also incorporates social constructivism into the social learning principle, explaining that, for most indidivuals, “learning is most effectively carried out via social interactions”. He also draws an interesting comparison between physics students and physics teachers, contrasting the need of students to learn in a social context with the desire or ability of most physicists to learn physics individually. The individuality principle explicitly addresses the differences in schema and learning styles between students, where Redish stresses that “any population of students will have a significant variation in a large number of cognitive variables”, challenging instructors to find the right environment to produce the best possible learning in the largest number of students.

As stated previously, the pedagogy known as interactive-engagement was defined by Hake as “designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on and hands-on activities which yield immediate feedback through discussion” (Hake 1998). Interactive-engagement courses, by definition, can address the principles of both Karplus’ Learning Cycle and Redish’s Cognitive Model. The efforts to improve physics education at UNC-CH made use of two primary interactive-engagement methodologies: SCALE-UP and the New Studio. As described in Chapter 1, SCALE-UP was primarily used during the initial implementation of interactive-engagement, and was taught concurrently with standard lecture based sections. The New Studio format was adopted as the teaching methodology as all physics classes at UNC-CH transitioned to interactive-engagement, and is referred to as Lecture/Studio within the context of UNC-CH’s efforts.

2.3.1 SCALE-UP

The goal of the Student-Centered Activities for Large-Enrollment University Physics (SCALE-UP) Project was to “establish a highly collaborative, hands-on, computer-rich, interactive learning environment in large-enrollment physics courses” (Beichner 1999).

SCALE-UP was developed at North Carolina State University, where thousands of students enroll in introductory physics classes each year. It utilized a cooperative learning approach, where students work through activities in groups of 3-4 students each. However, unlike other physics studio approaches, the program is designed to have student/faculty ratios of 25-50:1, allowing for classes of up to 120 students with 2-4 instructors (Beichner and Saul 2004).

There are often challenges inherent in implementing group work. Beichner defines two types of students that are often problematic: better students who avoid group work, and poorer students who can be lazy and refuse to participate. SCALE-UP addresses each of these differently. The course offers "teamsmanship points" to group members if a group's test average is above 80%, motivating the better students to assist their teammates. Students who are lazy can be fired from their group for poor performance, whereby they would be required to do an entire group's work alone. Groups are also encouraged to write contracts to manage their own group's operation (Beichner and Saul 2004).

The learning environment consists of round tables with comfortable chairs. Research was performed to determine the ideal table diameter; it was found that students preferred larger tables, but performed better in group work with tables approximately 7 feet in diameter. Each table seats three teams of three students each. The scheme that they use to identify each table and group provides a great deal of flexibility in breaking the class up to perform different tasks or activities. Rooms are equipped with a computer, a video overhead system, and multiple ceiling-mounted projectors. Each team has one laptop, and white boards are mounted on the walls to encourage collaborative problem solving while also allowing instructors to monitor group progress.

SCALE-UP's developers have created a variety of lesson plans to use during the course – some created especially for SCALE-UP, while others were modified from existing curricula. There are two types of group activities. Tangibles are quick tasks where hands-on measurement or observations are conducted by students. Ponderables require no observation, but require some sort of estimation or web search to determine a particular value or answer. Simulation packages and programs are used for students to experiment and better understand a particular concept. Because SCALE-UP incorporates so many group activities, labs can concentrate on topics beyond a general physics lab course, like experimental design, hypothesis testing, and error analysis.

In the 2004 paper, Beichner and Saul find that students in SCALE-UP have improved performance on final exam problems and concept inventories as compared to students in passive lecture courses (2004). With regard to the Force Concept Inventory, the normalized gain of SCALE-UP courses courses taught by Beichner from Spring 1997 until Fall of 2002 exceeded both the national FCI average of $\langle g \rangle = 0.22$, as well as Hake's moderate learning gain threshold of $\langle g \rangle = 0.3$. They also found that student's class attendance was improved, with over 90% of students present on average throughout a course.

2.3.2 New Studio

The New Studio format, also known as the Lecture/Studio format, was developed at Kansas State University (K-State) as the result of a department-wide effort to convert to an interactive-engagement physics education approach (Sorensen et al. 2006). In the traditional format, each lecture course at K-State consisted of approximately 150 students, and met for one hour twice per week (Churukian 2002). In addition, lecture courses were divided into recitation sections of 40 students, also meeting for one hour twice per week, and laboratory sections of 30 students, meeting for two hours once per week.

The K-State physics department had considered transitioning to the CUPLE Studio Physics format developed by Jack Wilson at Rensselaer Polytechnic Institute. In this format, large lectures are completely eliminated and replaced with a studio-only format. This concept was appealing, but the department was concerned that this method would increase the faculty teaching load and would require more classroom space than was available to the department (Sorensen et al. 2006). Instead, the university developed a modified version of this format, known as New Studio. In the New Studio, large lectures are retained, continuing to meet for one hour twice per week. Instead of separate recitations and labs, however, students meet in studio twice a week for two hours each.

Classrooms for the New Studio at K-State consist of ten tables for groups of up to four students. Each table is equipped with a computer. Each table also has access to a blackboard for the use of teams or instructors. The front of the room provides lecture tables for the instructors, an overhead projector, and a large blackboard. In the New Studio, lab demo activities take between 5 to 30 minutes, and are designed to be examples of physics concepts or homework problems. Rather than passively observing an instructor's demonstration, students are involved in the demonstrations. Problem solving is integrated into the format, with assigned problems integrated with lab demonstrations.

K-State measured the effectiveness of the New Studio using the Force Concept Inventory. Significant learning gains were identified, with increases in normalized gain up to a factor of 2.5. With the implementation of New Studio, all reported normalized gain from Spring 2000 to Spring 2001 well exceeded Hake's moderate gain threshold of $\langle g \rangle = 0.3$, and the investigators determined that the New Studio method "compare[d] well to other, innovative instructional methods recently reported," indicating that the combination of lecture and studio retained the learning benefits associated with courses taught wholly in an interactive-engagement format.

3 METHODOLOGY

Two primary methods of data collection were used in this research: written concept surveys and demographic data provided by the Office of the University Registrar. Written concept surveys (or concept inventories) are "research-based multiple choice assessment instruments used to test students' conceptual understanding of a topic" (Madsen et al. 2013). When these surveys are completed by students at the beginning and end of a course, they allow for a measurement of the students' conceptual gain. UNC-CH has been administering these tests to introductory physics students since the fall semester of 2007. For demographic data, the University Registrar provided information for students enrolled in introductory physics classes since Fall 2007. The combination of these two data sources allow for this project's primary research questions to be addressed.

3.1 Written Concept Surveys

3.1.1 Force Concept Inventory

The Force Concept Inventory (FCI), first published by David Hestenes, Malcolm Wells and Gregg Swachkamer in 1992, is a multiple-choice test instrument specifically designed to assess "a student's overall grasp of the Newtonian concept of force." The motivation for the FCI came from the realization that most students' common-sense ideas about physics are "incompatible with Newtonian concepts" (Hestenes et al. 1992). The authors of the FCI state that it has three primary applications: as a diagnostic tool, to evaluate instruction, and as a placement test (although the authors strongly discourage this particular use).

The FCI works particularly well as a diagnostic tool. Each question requires the student to choose between the correct Newtonian concept and the common-sense alternatives. The authors believe that incorrect answers are more valuable than the correct Newtonian choice, as they allow instructors to gain insight about their students' misconceptions about Newtonian concepts. The authors suggest that this usage be combined with interviews based on students' incorrect answers, which allow the instructor to then modify their instructional technique based on students' beliefs about physics.

The other recommended application of the FCI is to evaluate instruction – the application used in this research study. At the time of the FCI's publication, the authors suggest that "only the post-test score counts" to determine effective physics instruction. They make this assertion based on the uniformly low scores they encountered while delivering the FCI pre-test to high school and college students. They believe that truly effective instruction should always result in a high post-test score, so the pre-test/post-test difference is irrelevant, and likely skewed due to the uniformly low pre-test scores. However, as discussed in Chapter 2, Hake (1998) developed an additional metric for quantifying learning gains – the normalized gain. The normalized gain will be fully presented in the next chapter.

3.1.2 Conceptual Survey of Electricity and Magnetism

The Conceptual Survey of Electricity and Magnetism (CSEM), developed in 2001, is a multiple-choice test instrument designed to "assess students' knowledge about topics in electricity and magnetism" (Maloney et al. 2001). Similar to the FCI in that it can be used in a pre-test/post-test design, the CSEM differs in that it is considered to be a broad survey instrument, rather than focusing on one area – as the FCI does with respect to force in Newtonian mechanics. The authors acknowledge that the CSEM does not serve completely as a conceptual inventory in the same manner as the FCI. This is primarily due to both the breadth of topics contained within electricity and magnetism, as well as the limited literature available regarding students' misconceptions with respect to electricity and magnetism. Thus, some of the topics evaluated within the CSEM require a priori knowledge. Even with these limitations, the CSEM is considered to be both valid and reliable, and the authors believe that it can provide an "estimate of student learning for some of the more important ideas in electricity and magnetism" (Maloney et al. 2001).

3.1.3 Diagnostic Test Administration and Recordkeeping

There was a great deal of variation in the administration of the Force Concept Inventory and the Conceptual Survey of Electricity and Magnetism. Prior to the transition to Lecture/Studio, concept inventories were administered during the lab section by a teaching assistant. With regard to instructions, teaching assistants were either provided a specific script to read to students, or provided with general guidance. Faculty also delivered test instructions, although on an irregular basis (Churukian 2016a). After the transition to Lecture/Studio, concept inventories were administered during studio time, and teaching assistants are provided no specific instructions to read. In all cases, students were advised that they would receive participation credit for completing the survey. UNC-CH has FCI and CSEM data available for courses starting in Fall 2007. However, this thesis only used data beginning in Fall 2008 due to some apparent irregularities in the spreadsheets containing the data for the 2007-08 academic year.

Students are provided a paper copy of the appropriate concept inventory, and answer questions using a form on a departmental website. Test answers are downloaded from the website and entered into an Excel spreadsheet that performs grading and basic statistical analysis. For the purposes of this study, data were exported from Excel into a comma-separated-value format and imported into a statistical software package for analysis.

3.2 University Registrar Demographic Data

The University Registrar office provided student data in Microsoft Excel format for all courses covered under the study, based on final course rosters. The data provided in the dataset are summarized in Table 3.1. Each row of the dataset corresponded to one student's course enrollment during a single term. Students were only included in this set if they received a final grade in the course. Students who dropped the course after UNC-CH's drop date are included, as they receive an official grade of "W". Students entering in Fall of

Table 3.1: Description of database fields from registrar

Field	Name	Description
TERM	Term	Enrolled term
SUBJ	Subject	Subject code (PHYS)
CAT_NBR	Catalog Num.	Course number
SECT_NBR	Section Num.	Section number
DESCR	Course Description	Course name in catalog
ACAD_ORG	Academic Organization	Department (PHYS)
ACAD_CAREER	Academic Career	In this case, undergraduate (UGRD)
PROG	Student Program	Degree type (e.g., ASBS - Arts and Sciences, Bachelor of Science)
CLASS	Student class year	FR, SO, JR, SR
ACAD_PLAN	Academic Plan	Degree type + major (e.g., BSCHEMB - Bachelor of Science, Chemistry)
DESCR	Academic Plan	Plain text description of ACAD_PLAN
EMPLID	PID	Database internally refers to the PID as the EMPLID
NAME	Name	Student's name (Last, First Middle)
ACAD_GROUP	Academic Group	Usually CAS (College of Arts and Sciences)
UNT_TAKEN	Credits taken	Number of credits for course (4 in all cases)
CRSE_GRADE_OFF	Official course grade	Course grade
ACAD_PLAN_TYPE	Academic Plan Type	Type of academic plan (MAJ in almost all cases)
GENDER	Student gender	M or F
AFRAM	AFRAM flag	"AFRAM" if student identified as African American
AMIND	AMIND flag	"AMIND" if student identified as American Indian
ASIAN	ASIAN flag	"ASIAN" if student identified as Asian
ASPAC	ASPAC flag	"ASPAC" if student identified as an Asian-Pacific Islander
CAUCA	CAUCA flag	"CAUCA" flag if student identified as Caucasian
HSPLA	HSPLA flag	"HSPLA" flag if student identified as Hispanic/Latino
HWPAC	HWPAC flag	"HWPAC" flag if student identified as Hawaiian/Pacific Islander
NSPEC	NSPEC flag	"NSPEC" flag if student did not specify a ethnicity
OTHER	OTHER flag	"OTHER" flag if student identified as other
SINGLE_ETH	Single ethnicity field	Text description of student's ethnicity if they only identified ONE ethnicity, otherwise "Multi-Ethnic"

2014 had a drop period of ten days, while students entering prior to Fall of 2014 had eight weeks.

Raw data from the registrar's office were stored in a secured directory on the Physics and Astronomy Department file server. Working data were subsetted from the raw data by removing all identifying data except for the Person ID (PID). The PID is the University-approved identification number for students, faculty, and staff and is specific to the UNC-CH campus. The idea of hashing the PID for additional security was explored, however the University considers a PID to be public information and this procedure was eliminated from consideration. The working dataset also had all data fields removed that were not directly pertinent to the statistical analysis being performed (e.g., final course grade, major), and ethnicity data were reduced to the `SINGLE_ETH` field only. For importation into the statistical package, the dataset was converted into a comma-separated-value format.

3.3 Statistical Analysis

3.3.1 Regression Analysis

Regression analysis is a technique used to create a model that allows for the estimation of relationships between variables (Field et al. 2012). While there are a variety of regression models in use, this study uses the simplest: linear regression. In linear regression, the objective is to model observed data using a linear equation of the form

$$Y_i = (\beta_0 + \beta_1 X_i) + \varepsilon_i \tag{3.1}$$

where Y_i is the i-th observed dependent variable, X_i is the i-th case's score on the predictor variable, β_0 and β_1 are calculated coefficients of regression, and ε_i is the error in predicting Y_i from X_i, also known as the residual. In order to minimize ε_i, an ordinary least squares technique is applied, where the sum of squares of estimation is defined as

$$SS(E) = \sum_{i=1}^{N} \varepsilon_i^2 \tag{3.2}$$

and is combined with the previous equation and minimized with respect to β_0:

$$\frac{\partial SS(E)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i) \tag{3.3}$$

$$\frac{\partial SS(E)}{\partial \beta_0} = -2 \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{3.4}$$

and β_1:

$$\frac{\partial SS(E)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i) \tag{3.5}$$

$$\frac{\partial SS(E)}{\partial \beta_1} = -2 \sum_{i=1}^{N} x_i (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{3.6}$$

It can be shown that setting each partial derivative to zero and solving for each produces the equations that generate the regression coefficients in the bivariate case,

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \tag{3.7}$$

$$\beta_1 = \frac{\sum_{i=1}^{N} X_i Y_i - \frac{\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{N}}{\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}} \tag{3.8}$$

which allow for the completion of the model in Equation 3.1.

The quality of the model's fit can be evaluated by calculating another sum of squares: $SS(M)$, the model sum of squares, defined as

$$SS(M) = \sum_{i=1}^{N} (\hat{y}_i - \bar{Y})^2 \tag{3.9}$$

where $\bar{Y}$ is the average of the observed data, and $\hat{y}_i$ is the i-th predicted dependent variable based on the model. The ratio of the model sum of squares to the total sum of squares is a measure of the model's goodness-of-fit known as the coefficient of correlation, R^2:

$$R^2 = \frac{SS(M)}{SS(M) + SS(E)} \tag{3.10}$$

A more qualitative way to describe the coefficient of correlation is as the ratio of the variance explained by the model ($SS(M)$) to the total amount of variation in the sample ($SS(M) + SS(E) = SS(T)$). R^2 is generally interpreted as the percentage of the variation in the outcome that can be explained from the model, ranging from a perfect fit ($R^2 = 1$) to no correlation at all ($R^2 = 0$).

3.3.2 Analysis of Variance

Analysis of variance, or ANOVA, is a common statistical technique used to infer differences between the means of sample groups within an experiment. For example, a researcher may want to determine whether students in one of three different experimental groups (A_1, A_2 and A_3) really scored differently on a test, or whether the differences in scores were due to the random selection of the students. In this example, the null hypothesis – the assumption that there is no difference among groups – is that for each group A_i, all group means μ_i are comparable: $\mu_i = \mu$. This statistical technique is completely based on inference, and there is a probability that the observed result is not connected with the statistical analysis being performed. This probability is generally known as a p-value. If p is less than or equal to a threshold level, known as a significance level or α, then the null hypothesis may be rejected.

The null hypothesis in this case can be accepted or rejected by application of the F-test. In general, the F-test is a test of variance, and the test statistic is a ratio between estimates of the two main sources of experimental variance: the variance within a group versus the variance between groups, each calculated using a sum-of-squares technique. The between-group variance, or the model sum-of-squares $SS(M)$, is calculated in a similar fashion as with regression:

$$SS(M) = \sum_j n_j(\bar{Y}_{.j} - \bar{Y}_{..})^2 \tag{3.11}$$

where n_i is the number of observations in the j-th group, $\bar{Y}_{.j}$ is the mean of the j-th group, and $\bar{Y}_{..}$ is the mean of the full sample, also known as the grand mean. Note the similarities between this equation and Equation 3.9 – this model sum-of-squares is an extension of the previous equation to a scenario with multiple groups.

Similarly, the within-group variance, or the error sum-of-squares $SS(E)$, can be calculated as

$$SS(E) = \sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})^2 \tag{3.12}$$

where Y_{ij} is the i-th value of the j-th group and $\bar{Y}_{.j}$ is the mean of the j-th group. Then, the F-ratio can be defined as the ratio of the mean-square of the between-group variance and the mean-square of the within-group variance

$$F = \frac{MS(M)}{MS(E)} \tag{3.13}$$

$$F = \frac{SS(M)/(k-1)}{SS(E)/(N-k)} \tag{3.14}$$

where k is the number of groups and N is the total sample size. If the variance between groups dominates then F will increase, while if the variance within groups dominates F will decrease. If F exceeds a threshold value that is dependent on the number of degrees of freedom and the desired significance level, then the null hypothesis can be rejected allowing for the differences in group means to be considered statistically significant.

The example above is known as a one-way ANOVA, and can only be used for numeric data. This study utilized a two-way ANOVA, which allow for data to be split with regard to two independent variables, known as factors. The concept is the same as in the one-way case – variance between groups is compared to variance within groups to infer a statistically significant difference between groups.

3.3.3 Post hoc testing

After analysis of variance techniques indicate a statistically significant difference in group means, post hoc testing must be completed to determine the directionality of the difference, i.e., which means are significantly different from the others. ANOVA only allows for the inference of group mean differences, post hoc procedures indicate which groups differ. While a variety of different post hoc procedures exist, this project uses the Tukey Honestly Significant Difference test – known as the Tukey HSD test.

The Tukey HSD test is designed to test all means against each other pairwise: all means are tested against all other means. (Tabachnick and Fiddell 2007). To perform the procedure, calculate t_s for each pair of means:

$$t_s = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS(E)}{n_h}}} \tag{3.15}$$

where $\bar{X}_i$ and $\bar{X}_j$ are the i-th and j-th group means, $MS(E)$ is the mean square of the error, and n_h is the harmonic mean of the sample sizes of the groups A_i and A_j. The test statistic t_s is then compared against a critical value t_c, which is generally obtained by using a reference source. The critical value t_c is a function of the degrees of freedom of the sample's error term, the desired error rate, and the number of treatments in the analysis of variance. If $t_s > t_c$, then the means examined in the comparison have a statistically significant difference. The statistical package used for this project contains a function that can provide a specific p-value when provided with the results of an analysis of variance.

3.4 Statistical Software

This study utilized software developed by the R Project for Statistical Computing (R Core Team 2015). R is a freely-available software package that is similar to the S language and software that was developed at Bell Laboratories. Internal R subroutines handled all data handling and plotting, along with basic statistical tasks. Regressions and analysis of variance was performed using the Companion to Applied Regression (CAR) package (Fox and Weisberg 2011). While R provides internal functions that allow for a two-way analysis of variance, the functionality provided by the CAR package was more robust.

4 DATA AND ANALYSIS

In order to answer the research questions posed in Chapter 1, multiple-choice test instruments were used to survey students' understanding of various physics concepts. The results of these surveys were combined with demographic data provided by the Office of the University Registrar to understand the effectiveness of interactive learning, with and without consideration of race, gender, field of study, or final course grade.

4.1 Measuring learning gains

This study used two methods to quantify learning gains on the Force Concept Inventory (FCI) and the Conceptual Survey of Electricity and Magnetism (CSEM): the normalized gain $\langle g \rangle$ and the normalized change c.

4.1.1 Normalized gain

In a 6000-student survey of FCI scores from introductory physics courses, Richard Hake applied the "average normalized gain" to pre-test and post-test FCI data (Hake 1998). The average normalized gain $\langle g \rangle$, is defined as:

$$\langle g \rangle = \frac{\langle G \rangle}{\langle G \rangle_{maximum}} \tag{4.1}$$

$$\langle g \rangle = \frac{\langle \%\text{post-test} \rangle - \langle \%\text{pre-test} \rangle}{100\% - \langle \%\text{pre-test} \rangle} \tag{4.2}$$

where $\langle G \rangle$ is the average absolute gain, $\langle G \rangle_{maximum}$ is the average maximum gain, and $\langle \%\text{pre-test} \rangle$ and $\langle \%\text{post-test} \rangle$ are the average scores on the pre- and post-tests on a percentage basis. While Hake's particular application of this statistic was new, the normalized gain was originally used by Frank Gery in 1972 as a statistic to evaluate educational methods in economics. Due to its use by Hake, it is now commonly known as the "Hake gain" (Bao 2006). While the definition in Equation 4.2 is used to explicitly calculate $\langle g \rangle$, Equation 3.1 presents a clearer picture of what the normalized gain actually represents: the ratio of the average absolute gain $\langle G \rangle$ to the average maximum gain possible, $\langle G \rangle_{maximum}$. Hake considers $\langle g \rangle$ to be a "rough measure of the average effectiveness of a course in promoting conceptual understanding," and defined values of $\langle g \rangle$ consistent with either a lecture-based or interactive-engagement approach.

In his original paper, Hake only applied the normalized gain to the average pre- and post-instruction FCI scores of a course, rather than calculating an individual gain for each student. Continuing Hake's original application, most PER researchers continue to use the population average to determine the population gain for a particular course. However, some research has examined using the average of individual gains: negligible differences have been found between the two methods for courses of $N > 50$ (Bao 2006). There are several inherent issues with the calculation of individual gains – many of which are discussed in the next section. However, analysis of variance – the key statistical technique used in this study – requires the calculation of individual gains. Throughout the remainder of this chapter an individual gain and an average of individual gains will be denoted as g and $\bar{g}$, respectively. Any use of $\langle g \rangle$ implies the use of population averages and will be used for all other analysis.

4.1.2 Normalized change

This project also makes use of a newer metric known as the normalized change c. Developed by Jeffrey Marx and Karen Cummings, it corrects three perceived issues inherent in the use of normalized gain (Marx and Cummings 2007):

1. The normalized gain can generate a non-symmetric range of scores, ranging from any negative number to $+1$.

2. Normalized gain is biased toward low test scores – a student with a pre-test score of 20% can achieve a minimum normalized gain of -0.25, while a student with a pre-test

score of 80% can achieve a minimum normalized gain of -4.

3. Students achieving a pre-test score of 100% will have a normalized gain of $g = -\infty$ for all post-test scores.

To correct this issue, Marx and Cummings define the normalized change c as:

$$c = \begin{cases} \dfrac{\text{post} - \text{pre}}{100 - \text{pre}} & \text{post} > \text{pre} \\ \text{drop} & \text{post} = \text{pre} = 100 \text{ or } 0 \\ 0 & \text{post} = \text{pre} \\ \dfrac{\text{post} - \text{pre}}{\text{pre}} & \text{post} < \text{pre} \end{cases} \tag{4.3}$$

Equation 4.3 addresses the issues identified in the list above – scores are symmetric from -1 to $+1$, scores can be calculated for all students, and the effect of loss is proportional to the effect of gain[1]. From a more practical perspective, the use of c corrects situations where students did not take the post-test seriously, without resorting to dropping test scores as outliers – a process discussed in the next section.

4.2 Elimination of Outliers

As discussed above, the normalized gain has two primary behavior regimes: where $prescore < postscore$ and where $prescore > postscore$. The ability of g to go to $-\infty$ in the second regime produces a distribution that should be non-Gaussian in nature. To test the normality of the sample, the Shapiro-Wilk test was used. The Shapiro-Wilk test statistic W is defined as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\left(\sum_{i=1}^{n} x_i - \bar{x}\right)^2} \tag{4.4}$$

[1]Marx and Cummings present several examples of this proportionality in the paper describing normalized gain.

where x_i are the samples, $x_{(i)}$ are the i-th order statistics, $\bar{x}$ is the sample mean, and a_i is a coefficient determined from a Gaussian distribution and their covariance matrix. R provides a function, `shapiro.test`, which allows for the calculation of the Shapiro-Wilk test on a distribution for $N < 5000$. This function fails when applied to the UNC-CH FCI dataset, as $N > 5000$. However, if the distribution of the normalized gain is Gaussian, it would be expected to be independently Gaussian with respect to two random samples taken from the larger sample. In this case, the Shapiro-Wilk test provides two results, both suggesting that the distribution of normalized gain scores is non-Gaussian ($W_1 = 0.55353$, $p_1 < 0.001$; $W_2 = 0.32894$, $p_2 < 0.001$).

This is of concern when determining how to remove outliers from the data set. In general, scores may be outliers because a student's true performance deviated significantly from others in the course, or because a student did not complete the testing in good faith. Any removal of outliers should preserve the first and remove the second. It is difficult, however, to differentiate between the two. Again examining the UNC-CH FCI dataset, we find the two most extreme outliers to be $g = -23$ and $g = -21$. An examination of these two cases show that for the case of $g = -23$, the student scored a 29/30 on the pre-test and a 6/30 on the post-test. Similarly, for the case $g = -21$, the student scored a 29/30 on the pre-test and a 8/30 on the post-test. Just examining the raw decrease in test scores (23 and 21 questions) suggests the possibility that the students did not complete the post-test in good faith.

It is a common practice to consider data more than $\pm 3\sigma$ away from the mean of a distribution to be an outlier. However, this assumption is based on the data approximating a Gaussian distribution, and would be inappropriate in this case. However, there is no underlying statistical meaning behind the choice of $\pm 3\sigma$ other than the fact it is a round number and corresponds to an extremely low probability on a Gaussian distribution (two-tailed $p = 0.0027$). It should also be possible to choose a cut-off value of g that would include all reasonable scores, yet exclude those that would be highly improbable due to chance.

In order to set a cut-off value for g, the possible ways to achieve a variety of negative normalized gains were examined. Table 4.1 details how normalized gains of $g = -4$, $g = -5$ and $g = -6$ can be achieved. These possible scores were based on the total number of questions in the FCI. Pre-test scores of 30 were excluded as they produce an infinite result for any post-test score.

Table 4.1: Possible normalized gains by pre-test score

$g \leq -4$			$g \leq -5$			$g \leq -6$		
PRE	POST	Δ	PRE	POST	Δ	PRE	POST	Δ
29	25	-4	29	24	-5	29	23	-6
28	20	-8	28	18	-10	28	16	-12
27	15	-12	27	12	-15	27	9	-18
26	10	-16	26	6	-20	26	2	-24
25	5	-20	25	0	-25	25	—	—
24	0	-24	24	—	—	24	—	—

Consider the case of $g \leq -5$ in Table 4.1. In order to obtain this normalized gain or less on the FCI, a student could have a pre-test score of 29 and score below a 24 on the post-test, or a score of 28 on the pre-test and a post-test score of less than 18. This pattern continues until a pre-test score of 24 is reached – it is impossible to obtain $g \leq -5$ with a pre-test score of 24 or less. If a student scored a 28 or less on the post-test, they would have to have a total net loss of 10 questions to obtain $g \leq -5$. While this would be possible if the student were guessing, it's relatively unlikely for a student to obtain a true score of 28 on the pre-test and miss 10 additional questions on the post-test with a good-faith effort. The margins are narrower for students who score a 29 on the pre-test: it would only take the loss of five questions to result in $g \leq -5$, still somewhat unlikely for a student with a true pre-test score. Given these results, this study will remove any scores of $g \leq -5$ from the FCI and CSEM datasets during analysis. In order to analyze the robustness of c, the analysis of normalized change will include the entire dataset.

4.3 Force Concept Inventory Analysis

4.3.1 Outlier and Metric Analysis

In order to contrast the results of using g versus c, a regression was performed. As expected, the regression showed a statistically significant linear relationship ($F(1,7861) = 3670$, $p < 0.001$), yet resulted in a R^2 of only 0.3182. To compare, a regression was performed on g versus c with the elimination of all scores of $g \leq -5$. This resulted in a more robust linear relationship ($F(1,7836) = 1.61 \times 10^4$, $p < 0.001$) and correlation coefficient ($R^2 = 0.674$). Figure 4.1 presents the graphical results of these regressions.

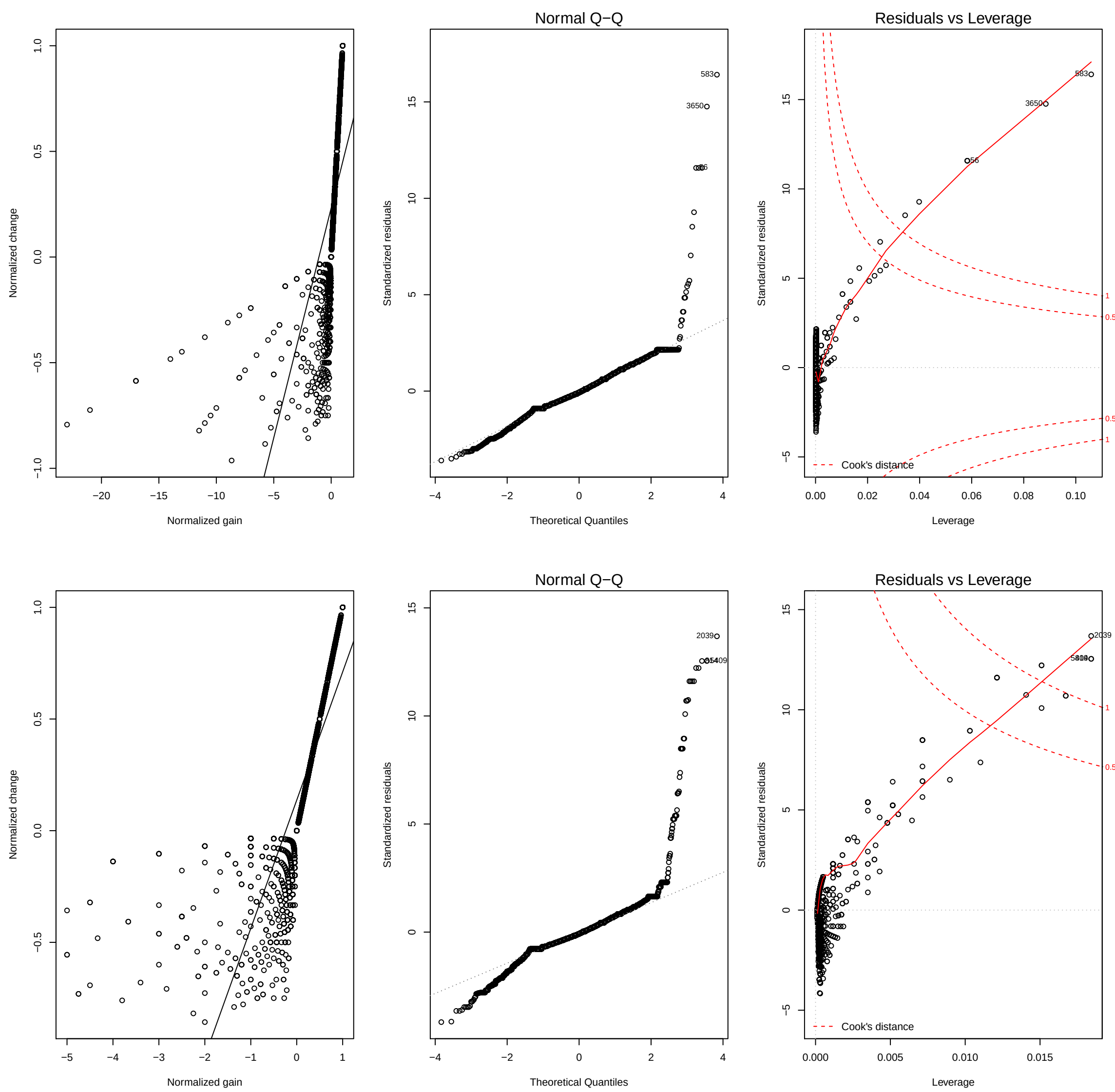
The graphical analysis also suggests an improved dataset with the removal of outliers: the first column of Figure 4.1 indicate fewer outliers in the dataset with outliers excluded, and greater linearity with respect to the regression line. The Q-Q plot measures the normality of the residuals from the regression – less divergence in the Q-Q plot line indicates greater linearity, and also demonstrates adherence to the ANOVA assumption of Gaussian residuals. The final column of Figure 4.1 contains a Residuals-Leverage plot, which is a diagnostic indicating influential cases that are often outliers. The Cook distance (D_i) is a quantitative measure of the influence of a particular data point, and values greater than 1 is considered a cut-off to determine influential data points. As the number of points exceeding D_i is reduced, this indicates a reduction in the number of outliers and their influence on the dataset. Based on these results, the analysis was continued with $g \leq -5$ excluded, resulting in a net loss of 25 cases (0.31%).

4.3.2 Demographics

The normalized change dataset consisted of 7863 total cases: 4118 female and 3745 male. Due to the exclusion of outliers, the normalized gain dataset was slightly smaller, 7838 total cases with 4114 female and 3724 male. Ethnicities between the two datasets are shown in Table 4.2, with the full dataset on the left and with outliers excluded on the right. Based on the number of students in each ethnic group, students were refactored into three

Figure 4.1: Regression results of g versus c, with and without identified outliers

The top row includes identified outliers where $g \leq -5$, the bottom row excludes these values. The first plot is a scatterplot of c versus g with a regression line overplot; the second plot is a standard Q-Q representation of the regression, and the third plot is a plot of residuals versus leverage.



groups to maintain statistical power: White, Asian and Other. This combination results in the breakdown shown in Table 4.3.

Table 4.2: Ethnicities in FCI datasets

Table (a) is the full dataset with outliers for use in the normalized change analysis. Table (b) is the dataset with outliers removed for use in the normalized gain analysis.

(a)

American Indian	26
Asian	1,295
Asian/Pacific Islander	44
African-American	522
Caucasian	4,827
Hawaiian	1
Hispanic/Latino	181
Multi-Ethnic	784
Not Specified	121
Other	62

(b)

American Indian	26
Asian	1,290
Asian/Pacific Islander	44
African-American	522
Caucasian	4,808
Hawaiian	1
Hispanic/Latino	181
Multi-Ethnic	783
Not Specified	121
Other	62

Table 4.3: Ethnicities in FCI datasets with combined ethnic factors

Table (a) is the full dataset with outliers for use in the normalized change analysis. Table (b) is the dataset with outliers removed for use in the normalized gain analysis.

(a)

Asian	1,295
White	4,827
Other	1,741

(b)

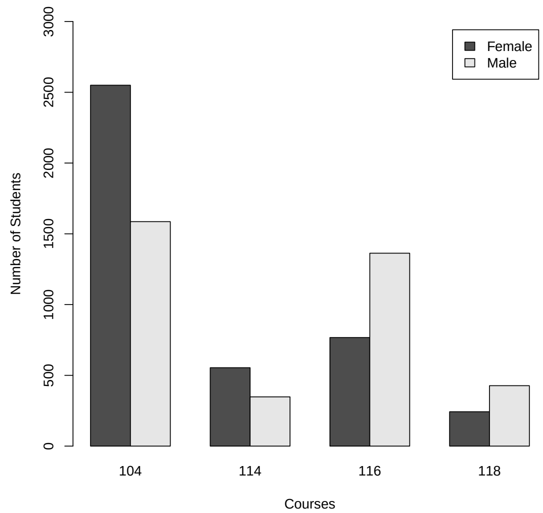
Asian	1,290
White	4,808
Other	1,740

Figure 4.2 shows total enrollment over the course of the study in each physics course by gender. As is common among college physics courses, enrollment in the IPLS (104/114) is dominated by female students, and (116/118) courses are composed of a majority of males. There was not enough data available to determine whether the ratio of male to female students was altered based on the new curriculum.

4.3.3 Learning gains

Learning gains were assessed using both the normalized gain and normalized change. To examine the normalized gain over the course of the study period, $\langle g \rangle$ was used, rather

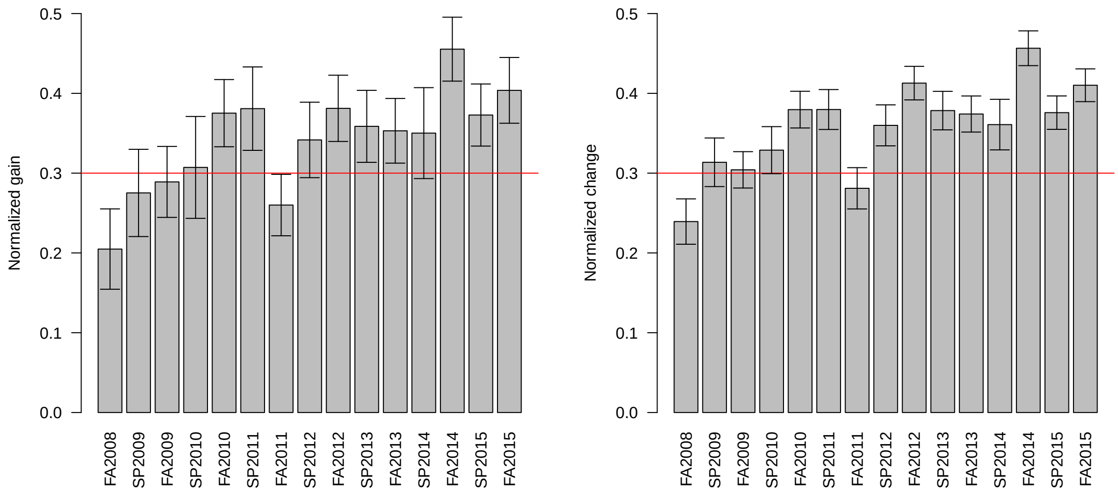
Figure 4.2: Enrollment in mechanics courses by gender



than the individual calculation of g for each student, preserving the ability to compare results with other studies. Normalized change was calculated individually for each student; results in this section are stated in terms of c_{ave}. Figure 4.3 presents learning gains by term in the IPPS course. Several trends are obvious: the normalized gain appears, in most cases, to trail the normalized change in terms of magnitude. This is most likely due to the fact that the normalized change is still restricted to the domain $-1 \leq c \leq 1$, while our elimination of outliers in the normalized gain only imposes a restriction of $-5 \leq g \leq 1$. There also appears to be an upward trend in both normalized gain and normalized change through all semesters, however there is some apparent variability in the normalized change between the fall and spring semesters, which correlates with the differing enrollment patterns for these terms. There is also an apparent jump in scores during the Fall semester of 2010, possibly correlating with the introduction of an interactive-engagement section in that term and all following terms, until the transition to full Lecture/Studio in Fall 2013. A possible confounding factor exists in this analysis, as the course mode changed exclusively to Lecture/Studio in Fall 2013, while the curriculum change occurred in Fall 2014. An overall upward trend was also noted in the results for the IPLS course (Figure 4.4).

Figure 4.3: FCI-measured learning gains by term, IPPS courses

Plot (a) is the learning gain as measured by the normalized gain using the dataset without outliers. Plot (b) is the learning gain as measured by the normalized change using the full dataset



(a) (b)

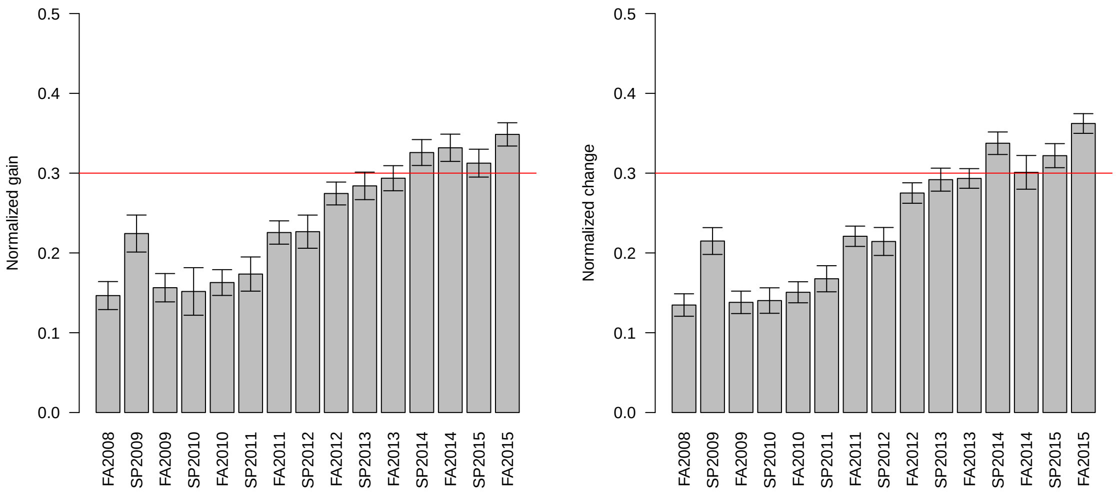
Interactive-engagement instruction for these courses only began during the Fall 2014 term, yet the upward trend in learning gains was consistent prior to Fall 2014. A slight decrease in c was noted for Fall 2014, likely due to challenges associated with the implementation of interactive-engagement courses – the decrease appears to have resolved itself during Spring 2015, and was not apparent in $\langle g \rangle$. There is also less variability between the normalized gain and normalized change, likely because the increased N associated with the IPLS courses lends less weight to the negative normalized gain scores.

While there are apparent trends suggesting enhanced learning gains with the adoption of interactive-engagement courses, quantitative results will require an analysis of variance (ANOVA). Student test scores were classified by their modality – `T` for enrollment in a traditional course and `I` for enrollment in an interactive-engagement course. Course numbers could not be used for this analysis as there was at least a section of interactive-engagement in every term of 116 from Fall 2010 until Spring 2014.

For the analysis of learning gains based on modality, two two-way ANOVAs were

Figure 4.4: FCI-measured learning gains by term, IPLS courses

Plot (a) is the learning gain as measured by the normalized gain using the dataset without outliers. Plot (b) is the learning gain as measured by the normalized change using the full dataset



(a) (b)

performed for each course type using the `lm` and `Anova` commands in R – one considering normalized gain versus modality and another considering normalized change versus modality. Because of the unbalanced design (more students are present in the `T` sample versus the `I` sample), a Type III sum-of-squares was used. In this type of calculation, the sum-of-squares is calculated as if each effect were listed last in the model, resulting in an orthogonal comparison. This same procedure was used for all ANOVA calculations in this thesis. For IPPS courses, both normalized gain ($F(1,2798) = 22.945$, $p < 0.01$) and normalized change ($F(1,2817) = 47.071$, $p < 0.01$) are statistically significant. Similarly, for IPLS courses, the normalized gain ($F(1,5036) = 70.498$, $p < 0.001$) and normalized change ($F(1,5042) = 112.58$, $p < 0.001$) are statistically significant. These findings indicates that the course modality does have a statistically significant impact on learning gains as measured by the Force Concept Inventory. Because this analysis was limited to one independent variable (modality) and two factors (traditional versus interactive-engagement), a post hoc test was not required, and the statistical analysis shows that implementation of interactive-engagement has led to increased learning gains, regardless of instructor.

4.3.4 FCI learning gains and gender

Next, learning gains were examined as a function of both gender and modality. Table 4.4 shows the learning gains by gender and modality for courses in the IPPS sequence. There is an overall increase in $\langle g \rangle$ and c for males and female students in interactive-engagement courses compared to traditional courses. There is also a decrease in Δ_{M-F}, the male-female learning gap, when transitioning from traditional to interactive-engagement courses for both $\langle g \rangle$ and c. Figure 4.5 indicates that both males and females enrolled in interactive-engagement courses exceeded the $\langle g \rangle = 0.3$ and $c = 0.3$ threshold defined by Hake as moderate-gain.

Table 4.4: FCI-measured learning gains by gender and modality, IPPS

Table (a) is the normalized gain analysis without outliers. Table (b) is the normalized change analysis using the full dataset.

(a)

	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
Male	0.33 ± 0.02	0.41 ± 0.02
Female	0.27 ± 0.02	0.37 ± 0.02
Δ_{M-F}	0.06 ± 0.02	0.04 ± 0.02

(b)

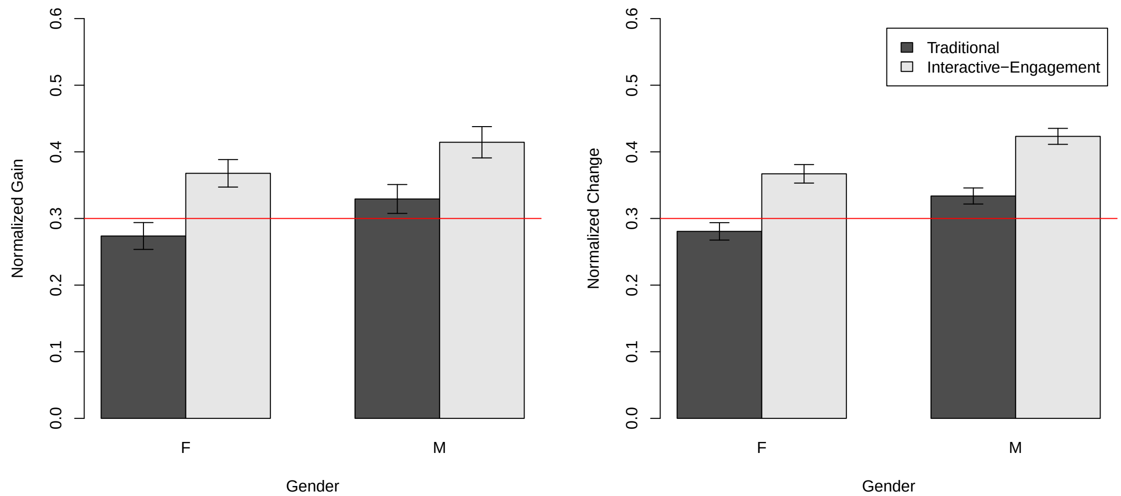
	c_T	c_{IE}
Male	0.33 ± 0.01	0.42 ± 0.01
Female	0.28 ± 0.01	0.37 ± 0.01
Δ_{M-F}	0.05 ± 0.01	0.04 ± 0.01

For the IPLS sequence (Table 4.5), there was an overall increase in $\langle g \rangle$ and c for males and females, although Δ_{M-F} also increased for both. Figure 4.6 shows that male students exceeded the $\langle g \rangle = 0.3$ and $c = 0.3$ moderate-gain threshold for interactive-engagement courses in the IPLS sequence, while female students either met or were within 3% of the threshold when considering the margin of error.

Two-way ANOVAs were performed with $\langle g \rangle$ and c as dependent variables. For each ANOVA, gender and course modality were examined as independent factors. For the difference in normalized gain in IPPS courses, the differences in means were found to be significant for only the main effect of modality ($F(1,2796) = 22.35$, $p < 0.001$). The main effect of gender with normalized gain was found to be non-significant ($F(1,2796) = 0.51$,

Figure 4.5: FCI-measured learning gains by gender and modality, IPPS

Plot (a) is the learning gain as measured by the normalized gain using the dataset without outliers. Plot (b) is the learning gain as measured by the normalized change using the full dataset



(a) (b)

Table 4.5: FCI-measured earning gains by gender and modality, IPLS

Table (a) is the normalized gain analysis without outliers. Table (b) is the normalized change analysis using the full dataset.

(a)

	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
Male	0.25 ± 0.01	0.39 ± 0.02
Female	0.212 ± 0.005	0.30 ± 0.01
Δ_{M-F}	0.04 ± 0.01	0.09 ± 0.02

(b)

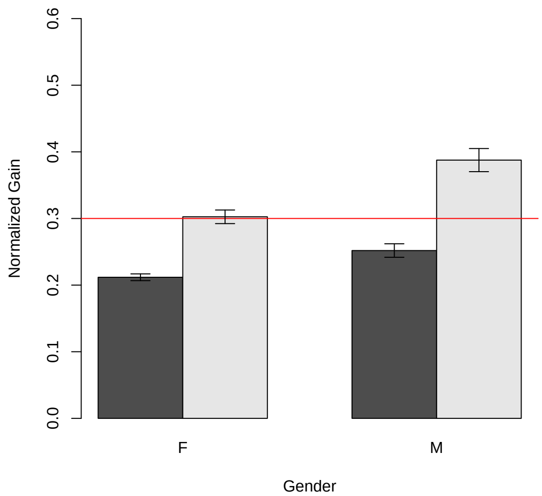
	c_T	c_{IE}
Male	0.25 ± 0.01	0.38 ± 0.01
Female	0.20 ± 0.01	0.29 ± 0.01
Δ_{M-F}	0.05 ± 0.01	0.09 ± 0.01

$p = 0.473$). Interaction between gender and course modality was not observed. Post hoc testing was not required as each factor had only two levels.

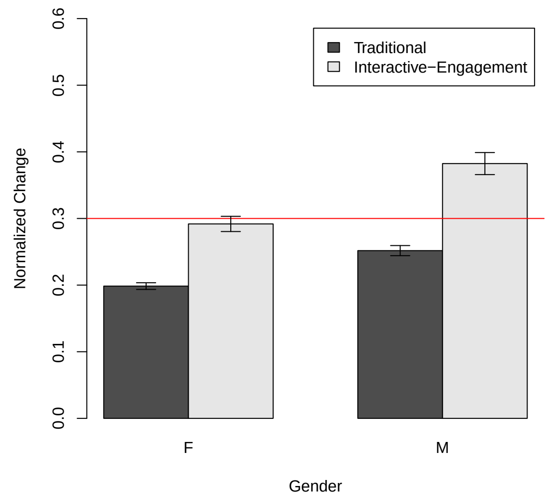
Analysis of the normalized change in the IPPS sequence also found a statistically significant difference in means with regard to course modality ($F(1,2796) = 36.71$, $p < 0.001$), but also found a statistically significant difference with regard to gender ($F(1,2796) = 20.35$, $p < 0.001$). No interaction between gender and course modality was observed. These statistically-significant result supports the identified increase in mean learning gains

Figure 4.6: FCI-measured learning gains by gender and modality, IPLS

Plot (a) is the learning gain as measured by the normalized gain using the dataset without outliers. Plot (b) is the learning gain as measured by the normalized change using the full dataset



(a)



(b)

and the reduction in the observed gender gap shown in Table 4.4 and Figure 4.5, indicating that the introduction of interactive-engagement resulted in an increase in learning gains across genders, as well as a decrease in the observed gap in performance between male and female students in IPPS courses.

In the case of normalized gain in the IPLS sequence, statistically significant differences were identified for both of the main effects: gender ($F(1,5034) = 77.132$, $p < 0.001$) and modality ($F(1,5034) = 20.354$, $p < 0.001$). There was also a statistically significant interaction between gender and modality ($F = 77.132$, $p < 0.001$). Existence of the interaction term precludes any further independent analysis of the main effect, so post hoc testing was not conducted.

Analysis of the normalized change of the IPLS sequence found statistically significant differences in means with regard to course modality ($F(1,5034) = 125.3$, $p < 0.001$) and gender ($F(1,5034) = 51.08$, $p < 0.001$). No significant interaction was observed between the two factors. These results for the normalized gain and normalized change support the increased mean scores in IPLS courses shown in Table 4.5 and Figure 4.6, indicating that the implementation of interactive-engagement resulted in an increase in learning gains across genders in IPLS courses.

4.3.5 FCI learning gains and ethnicity

As stated earlier, ethnicities as provided by the Office of the University Registrar were refactored into three groups: White, Asian, and Other. Normalized gains and normalized changes are shown in Table 4.6 between traditional and interactive-engagement courses in the IPPS sequence. Increases in $\langle g \rangle$ and c were observed for all ethnicities between traditional and interactive-engagement courses. There were also consistently observed decreases in the performance gap between White and Asian students (Δ_{W-A}) and White and Other students (Δ_{W-O}) when transitioning to interactive-engagement courses. Figure 4.7 illustrates that all groups exceeded the Hake moderate-gain threshold after implementation of interactive-engagement courses.

Table 4.6: FCI-measured learning gains by ethnicity and modality, IPPS

Table (a) is the normalized gain analysis without outliers. Table (b) is the normalized change analysis using the full dataset.

(a)

	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
White	0.34 ± 0.02	0.41 ± 0.02
Asian	0.27 ± 0.05	0.36 ± 0.02
Other	0.23 ± 0.04	0.37 ± 0.02
Δ_{W-A}	0.07 ± 0.05	0.05 ± 0.02
Δ_{W-O}	0.11 ± 0.04	0.04 ± 0.02
Δ_{A-O}	0.04 ± 0.06	-0.01 ± 0.06

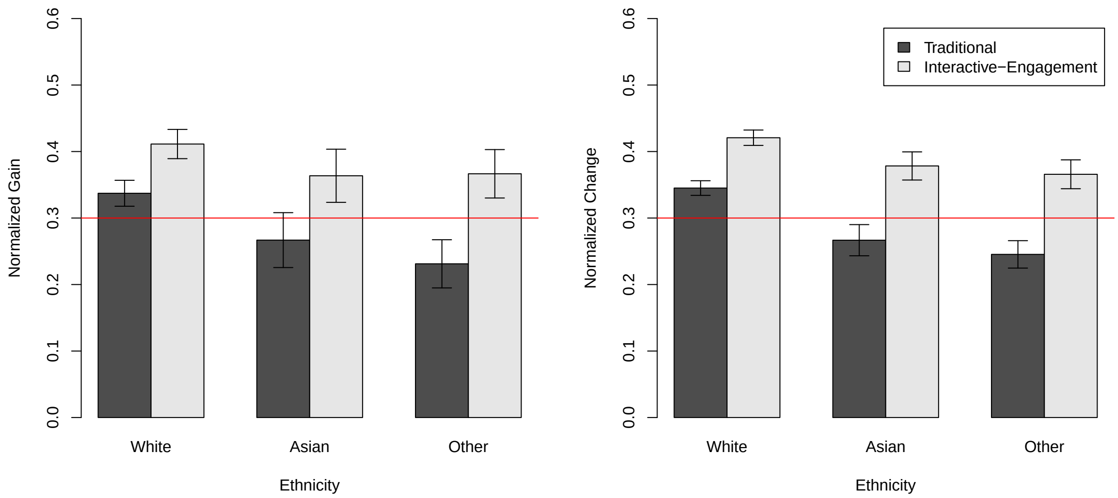
(b)

	c_T	c_{IE}
White	0.35 ± 0.01	0.41 ± 0.02
Asian	0.27 ± 0.02	0.38 ± 0.02
Other	0.25 ± 0.02	0.37 ± 0.02
Δ_{W-A}	0.08 ± 0.02	0.03 ± 0.02
Δ_{W-O}	0.10 ± 0.03	0.04 ± 0.02
Δ_{A-O}	0.02 ± 0.04	0.01 ± 0.03

In the IPLS sequence, learning gains were observed across all ethnicities for both the normalized gain and normalized change (Table 4.7). Performance gaps stayed constant or widened between White students (Δ_{W-A} and Δ_{W-O}), even with the consistent increase across all ethnic factors. In Figure 4.8, there was a significant difference in ethnic factors

Figure 4.7: FCI-measured learning gains by ethnicity and modality, IPPS

Plot (a) is the learning gain as measured by the normalized gain using the dataset without outliers. Plot (b) is the learning gain as measured by the normalized change using the full dataset



(a) (b)

and the Hake moderate-gain threshold; both White and Asian students exceeded the threshold for interactive-engagement courses when considering normalized gain and normalized change.

Table 4.7: FCI-measured learning gains by ethnicity and modality, IPLS

Table (a) is the normalized gain analysis without outliers. Table (b) is the normalized change analysis using the full dataset.

(a)

	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
White	0.24 ± 0.01	0.36 ± 0.01
Asian	0.21 ± 0.01	0.31 ± 0.03
Other	0.19 ± 0.01	0.29 ± 0.02
Δ_{W-A}	0.03 ± 0.01	0.05 ± 0.03
Δ_{W-O}	0.05 ± 0.01	0.07 ± 0.03
Δ_{A-O}	0.03 ± 0.01	0.02 ± 0.04

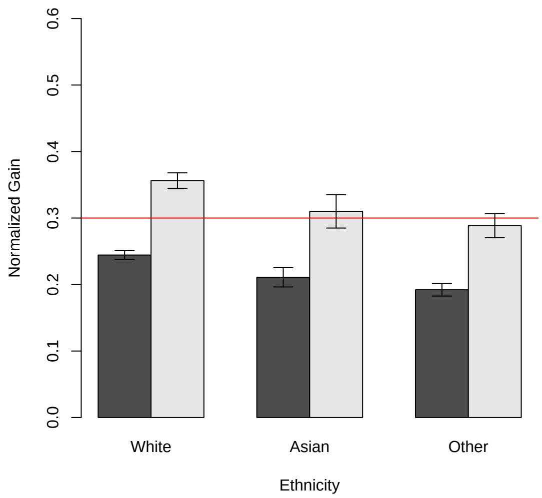
(b)

	c_T	c_{IE}
White	0.24 ± 0.01	0.35 ± 0.01
Asian	0.20 ± 0.01	0.31 ± 0.02
Other	0.18 ± 0.01	0.27 ± 0.02
Δ_{W-A}	0.04 ± 0.01	0.04 ± 0.02
Δ_{W-O}	0.06 ± 0.01	0.08 ± 0.03
Δ_{A-O}	0.02 ± 0.01	0.04 ± 0.04

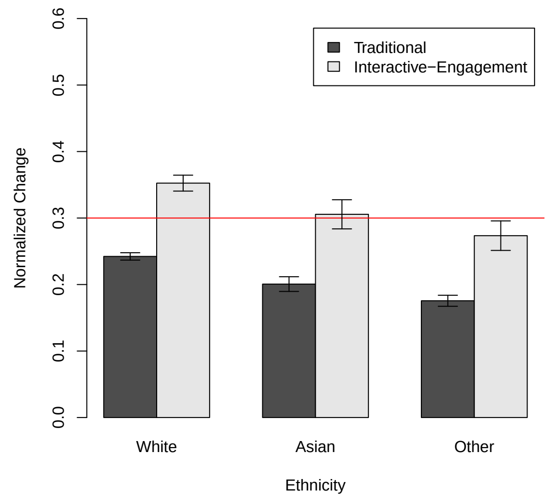
ANOVAs were conducted across course types (IPLS versus IPPS) and gain metric (normalized gain versus normalized change). In each, the gain metric served as the dependent variable, while the modality and ethnicity factors served as the independent variables.

Figure 4.8: FCI-measured learning gains by ethnicity and modality, IPLS

Plot (a) is the learning gain as measured by the normalized gain using the dataset without outliers. Plot (b) is the learning gain as measured by the normalized change using the full dataset



(a)



(b)

When examining the normalized gain in IPPS courses, statistically significant differences in means were found for both main effects: modality ($F(1,2794) = 26.46$, $p < 0.001$) and ethnicity ($F(2,2794) = 6.84$, $p < 0.01$). Tukey HSD post hoc analysis found statistically significant differences between the performance of White and Asian students and White and Other students ($p < 0.05$), while no statistically significant difference was found between Asian and Other students.

Similarly, ANOVA analysis of the IPPS course with respect to the normalized change found a statistically significant difference in means when considering both modality ($F(1,2794) = 38.488$, $p < 0.001$) and ethnicity ($F(2,2794) = 15.671$, $p < 0.001$). Post hoc analysis found similar results to the normalized gain ANOVA: statistically significant differences between the performance of White and Asian students and White and Other students ($p < 0.05$), while no statistically significant difference was found between Asian and Other students. These statistically significant differences with regard to the IPPS courses support the increased learning gains identified in Table 4.6 and Figure 4.7, and demonstrate that the implementation of interactive-engagement has resulted in increased learning gains and has reduced performance gaps with regard to ethnicity.

For the IPLS course, ANOVA analysis of the normalized gain found statistically significant differences in modality ($F(1,5032) = 49.167$, $p < 0.001$) and ethnicity ($F(2,5032) = 15.404$, $p < 0.001$), with no observed interaction effect. Tukey HSD post hoc analysis found statistically significant differences between the performance of White and Asian students and White and Other students ($p < 0.05$), while no statistically significant difference was found between Asian and Other students. Similar results were observed for the normalized change with modality ($F(1,5032) = 89.615$, $p < 0.001$) and ethnicity ($F(2,5032) = 17.871$, $p < 0.001$), with identical findings from post hoc testing. As with the IPPS courses, these statistically significant differences identified in the IPLS courses support the overall increase in learning gains shown in Table 4.7 and Figure 4.8, demonstrating that the implementation of interactive-engagement has resulted in increased learning gains across all ethnicities.

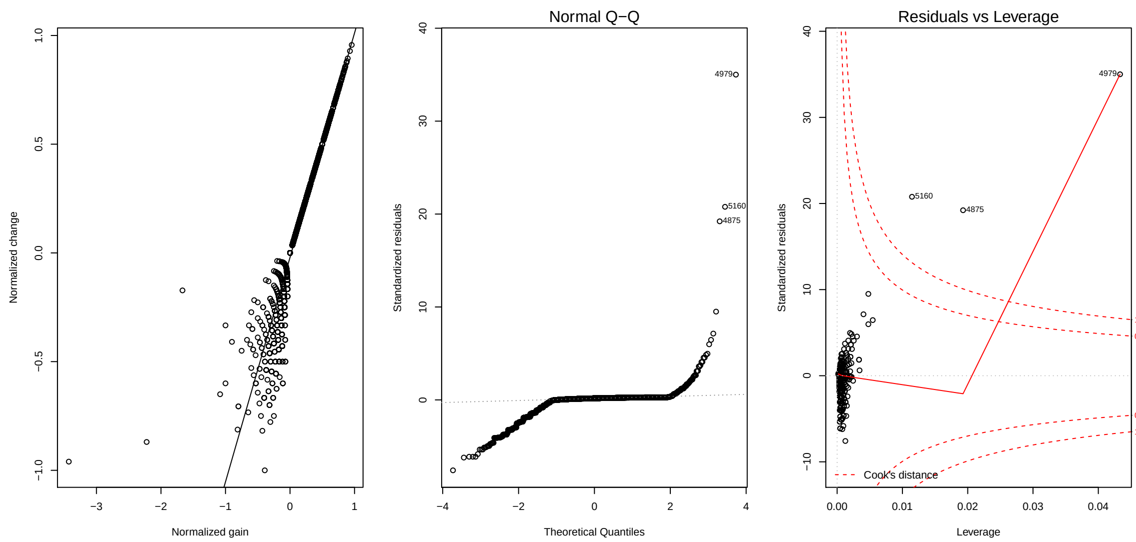
4.4 Conceptual Survey of Electricity and Magnetism Analysis

4.4.1 Outlier and Metric Analysis

In order to contrast the results of using g versus c, a regression was performed. As expected, the regression showed a statistically significant linear relationship ($F(1,5220) = 5.56 \times 10^4$, $p < 0.001$), and a correlation coefficient of ($R^2 = 0.914$). No cases met the outlier threshold of $g \leq -5$. The interpretation of Figure 4.9 is equivalent to that in Figure 4.1, and a detailed description of each plot can be found there. The first plot indicates a much higher degree of linearity between the normalized gain and normalized change. This is likely due to the fact that students are much less likely to get a high CSEM score, which limits the range of negative scores that can be obtained. In this case, the Q-Q plot still indicates a measure of normality in the residuals, as most of the points fall along the dashed line. Finally, there are three influential cases in the dataset with a $D_i > 1$. Further

Figure 4.9: Regression results of g versus c, with and without identified outliers

The first plot is a scatterplot of c versus g with a regression line overplot; the second plot is a standard Q-Q representation of the regression, and the third plot is a plot of residuals versus leverage.



analysis could lead to the removal of these cases, but this author wanted to maintain the outlier criteria established for the FCI.

4.4.2 Demographics

The normalized change dataset consisted of 5222 total cases: 2749 female and 2743 male. Because no outliers were excluded, the normalized gain dataset was identical. Ethnicities were also identical for both datasets, as shown below in Table 4.8. Based on the number of students in each ethnic group, students were refactored into three groups to maintain statistical power: White, Asian and Other. This combination results in the breakdown shown in Table 4.9. Figure 4.10 shows total enrollment over the course of the study in each physics course by gender. Similar to the findings in the FCI dataset, enrollment in the IPLS (105/115) is dominated by female students, while the IPPS (117/119) courses are composed of a majority of males. There was not enough data available to determine whether the ratio of male to female students was altered based on the new curriculum.

Table 4.8: Ethnicities in CSEM datasets

American Indian	17
Asian	814
Asian/Pacific Islander	42
African-American	284
Caucasian	3,298
Hawaiian	1
Hispanic/Latino	117
Multi-Ethnic	531
Not Specified	74
Other	44

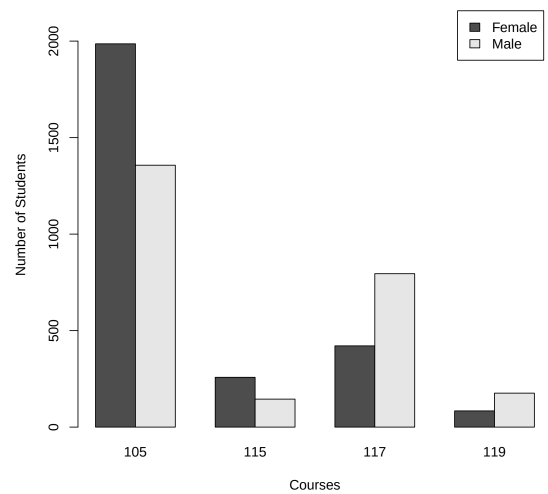
Table 4.9: Ethnicities in CSEM datasets with combined ethnic factors

Asian	814
White	3,298
Other	1,110

4.4.3 Learning gains

Learning gains were assessed using both the normalized gain and normalized change. To examine the normalized gain over the course of the study period, $\langle g \rangle$ was used, rather than the individual calculation of g for each student, preserving the ability to compare results with other studies. Normalized change was calculated individually for each student; results in this section are stated in terms of c_{ave}. Figure 4.11 presents learning gains by term in the IPPS courses. There is much less variability between the normalized gain and normalized change metrics. A slight upward trend is noted from Spring 2009 through Spring 2012. Similar trends were also noted in the results for the IPLS courses (Figure 4.12). With regard to IPLS courses, both normalized gain and normalized change scores are well below the scores for the IPPS courses. There is a large jump in both normalized gain and normalized change for the Spring 2015 term, when the interactive-engagement methodology was introduced. Although gains lessened during the Fall 2015 term, the scores during this term were still well above the thirteen previous terms prior to the implementation of

Figure 4.10: Enrollment in E&M courses by gender



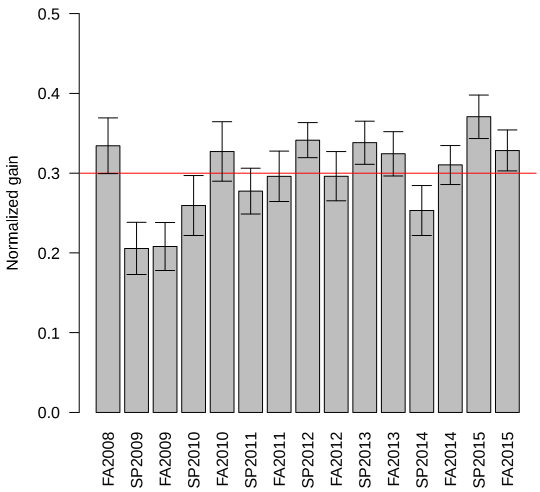
interactive-engagement learning.

As with the FCI dataset, analysis of variance was used to quantitatively examine the differences in means between the traditional and interactive-engagement courses. Student test scores were classified by their modality – `T` for enrollment in a traditional course and `I` for enrollment in an interactive-engagement course.

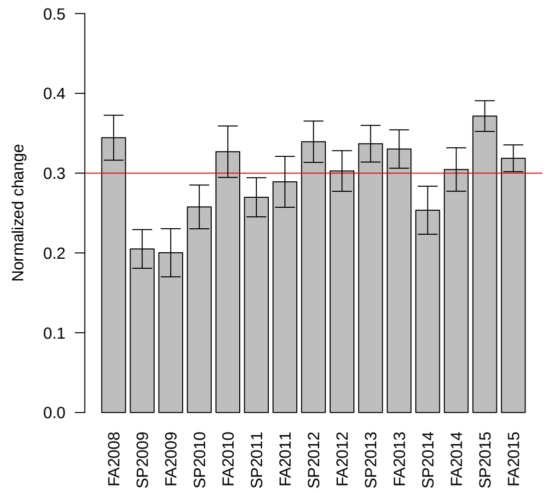
For the analysis of learning gains based on modality, two two-way ANOVAs were performed – one considering normalized gain versus modality and another considering normalized change versus modality. For the IPPS courses, both normalized gain ($F(1,1474) = 11.21$, $p < 0.001$) and normalized change ($F(1,1474) = 19.37$, $p < 0.001$) are statistically significant. A similar finding was noted for IPLS courses with regard to the normalized gain ($F(1,3744) = 35.41$, $p < 0.001$) and normalized change ($F(1,3744) = 27.34$, $p < 0.001$). These findings indicate that the course modality does have a statistically significant impact on learning gains as measured by the CSEM. Because this analysis was limited to one independent variable (modality) and two factors (traditional versus interactive-engagement), a post hoc test was not required. However, this statistically significant result supports the upward trends identified in Figure 4.11 and 4.12 associated with the implementation of

Figure 4.11: CSEM-measured learning gains by term, IPPS

Plot (a) is the learning gain as measured by the normalized gain. Plot (b) is the learning gain as measured by the normalized change.



(a)



(b)

interactive-engagement in introductory physics courses.

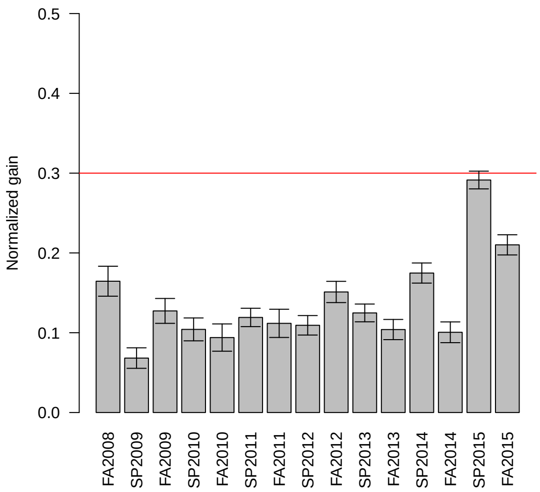
4.4.4 CSEM learning gains and gender

Next, learning gains were examined as a function of both gender and modality. Table 4.10 shows the learning gains by gender and modality for courses in the calculus sequence. There is an overall increase in $\langle g \rangle$ and c for males and female students in interactive-engagement courses compared to traditional courses. There is also a significant decrease in Δ_{M-F}, the male-female learning gap when transitioning from traditional to interactive-engagement courses for both $\langle g \rangle$ and c. Figure 4.13 indicates that only males enrolled in interactive-engagement courses exceeded the Hake moderate-gain threshold of $\langle g \rangle = 0.3$ and $c = 0.3$. Female students were within the margin of error of the threshold for both normalized gain and normalized change.

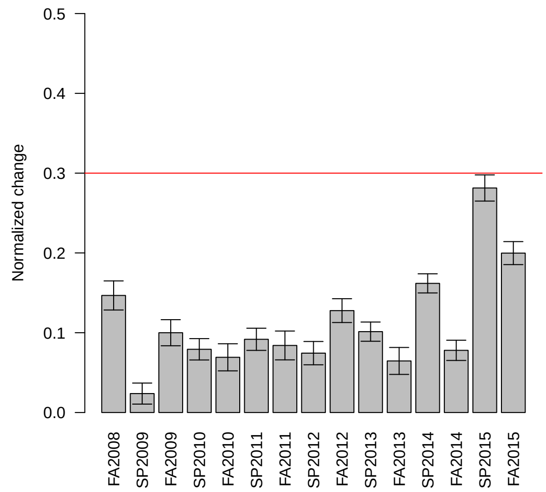
For the IPLS sequence (Table 4.11), there was an overall increase in $\langle g \rangle$ and c for males and females, although Δ_{M-F} also increased for both. No students met or exceeded the moderate gain threshold after the department's transition to interactive-engagement courses

Figure 4.12: CSEM-measured learning gains by term, IPLS

Plot (a) is the learning gain as measured by the normalized gain. Plot (b) is the learning gain as measured by the normalized change.



(a)



(b)

Table 4.10: CSEM-measured learning gains by gender and modality, IPPS

Table (a) is the normalized gain analysis. Table (b) is the normalized change analysis.

(a)

	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
Male	0.31 ± 0.01	0.35 ± 0.02
Female	0.20 ± 0.01	0.29 ± 0.02
Δ_{M-F}	0.11 ± 0.02	0.06 ± 0.03

(b)

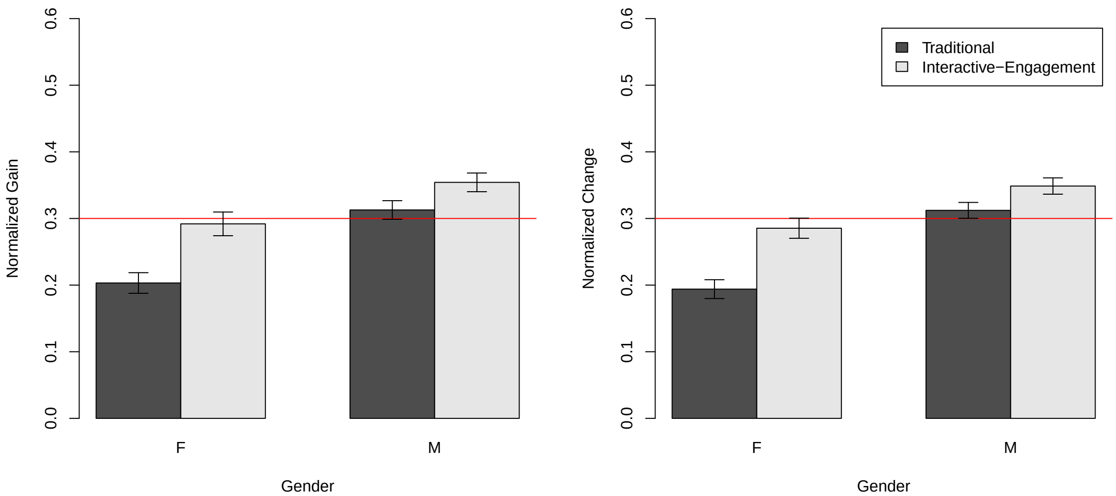
	c_T	c_{IE}
Male	0.31 ± 0.01	0.35 ± 0.01
Female	0.19 ± 0.01	0.29 ± 0.01
Δ_{M-F}	0.12 ± 0.02	0.06 ± 0.02

(Figure 4.14).

Two-way ANOVAs were performed with $\langle g \rangle$ and c as dependent variables. For each ANOVA, gender and course modality were examined as independent factors. For the difference in normalized gain in IPPS courses, the differences in means were found to be significant for both modality ($F(1,1472) = 12.44$, $p < 0.001$) and gender ($F(1,1472) = 26.74$, $p < 0.001$). Interaction between gender and course modality was not observed. Post hoc testing was not required as each factor had only two levels. Analysis of the normalized change in the IPPS sequence also found a statistically significant difference in means with regard to course modality ($F(1,1472) = 20.33$, $p < 0.001$) and gender ($F(1,1472) = 40.94$,

Figure 4.13: CSEM-measured learning gains by gender and modality, IPPS

Plot (a) is the learning gain as measured by the normalized gain. Plot (b) is the learning gain as measured by the normalized change.



(a) (b)

Table 4.11: CSEM-measured learning gains by gender and modality, IPLS

Table (a) is the normalized gain analysis. Table (b) is the normalized change analysis.

(a)

	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
Male	0.15 ± 0.01	0.22 ± 0.01
Female	0.10 ± 0.01	0.16 ± 0.01
Δ_{M-F}	0.05 ± 0.01	0.06 ± 0.02

(b)

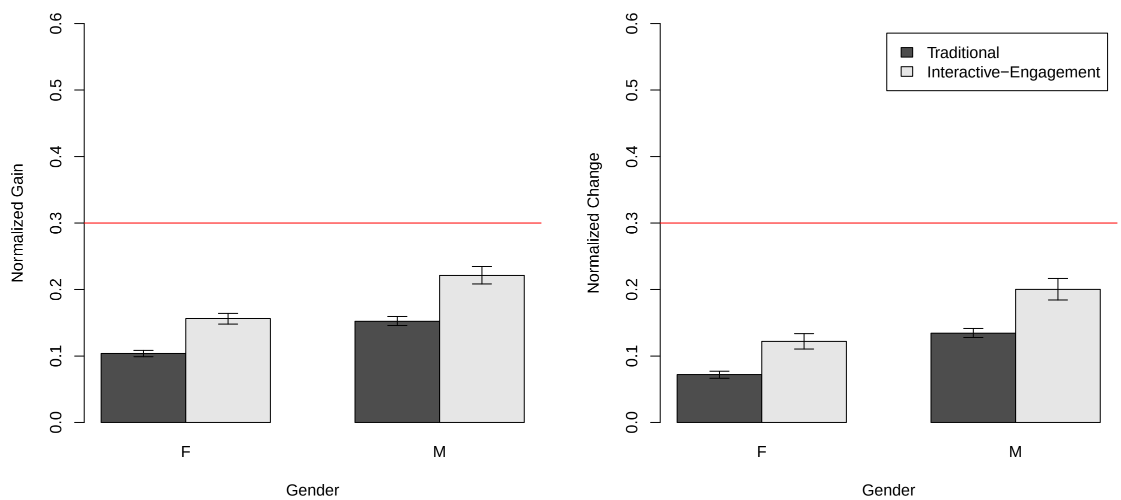
	c_T	c_{IE}
Male	0.13 ± 0.01	0.20 ± 0.02
Female	0.07 ± 0.01	0.12 ± 0.01
Δ_{M-F}	0.06 ± 0.01	0.08 ± 0.02

$p < 0.001$). No interaction between gender and course modality was observed. The statistically significant differences in means support the findings in Table 4.10 and Figure 4.13 of an overall increase in normalized gain and normalized change associated with the implementation of interactive-engagement learning in IPPS courses as well as a decrease in the performance gap betwen male and female students in IPPS courses.

For the IPLS sequence, statistically significant differences were identified in the normalized gain for both of the main effects: gender ($F(1,3742) = 53.59$, $p < 0.001$) and modality ($F(1,3742) = 41.79$, $p < 0.001$). No interaction was observed between the two factors. With respect to the normalized change, a statistically significant result was found

Figure 4.14: CSEM-measured learning gains by gender and modality, IPLS

Plot (a) is the learning gain as measured by the normalized gain. Plot (b) is the learning gain as measured by the normalized change.



(a) (b)

with course modality ($F(1,3742) = 31.85$, $p < 0.001$) and gender ($F(1,3742) = 47.00$, $p < 0.001$). No significant interaction was observed between the two factors. The statistically significant differences in means support the findings in Table 4.11 and Figure 4.14 of an overall increase in normalized gain and normalized change associated with the implementation of interactive-engagement learning in IPLS courses. The statistically significant finding also supports the observed increase in the performance gap betwen male and female students in IPLS courses, although this effect is small.

4.4.5 CSEM learning gains and ethnicity

As with the FCI analysis, ethnicities as provided by the Office of the University Registrar were refactored into three groups: White, Asian and Other. Normalized gains and normalized changes are shown in Table 4.12 between traditional and interactive-engagement courses in the IPPS sequence. Increases in $\langle g \rangle$ and c were observed for all ethnicities between traditional and interactive-engagement courses. Decreases in the performance gap between White and Asian students (Δ_{W-A}) and White and Other students (Δ_{W-O}) were not observed. Figure 4.15 illustrates that only the White students exceeded the Hake moderate-gain threshold after implementation of interactive-engagement courses. However, Asian and Other students were well within the margin of error when considering both normalized gain and normalized change.

Table 4.12: CSEM-measured learning gains by ethnicity and modality, IPPS

Table (a) is the normalized gain analysis. Table (b) is the normalized change analysis.

(a)

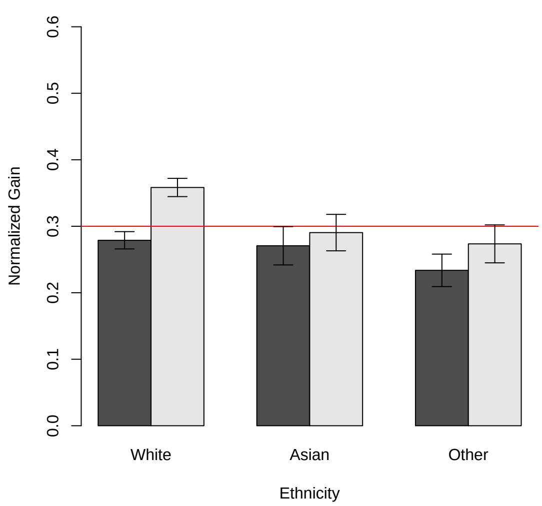
	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
White	0.28 ± 0.01	0.36 ± 0.02
Asian	0.27 ± 0.02	0.29 ± 0.02
Other	0.23 ± 0.02	0.27 ± 0.02
Δ_{W-A}	0.01 ± 0.02	0.07 ± 0.03
Δ_{W-O}	0.02 ± 0.02	0.09 ± 0.03
Δ_{A-O}	0.02 ± 0.02	0.02 ± 0.04

(b)

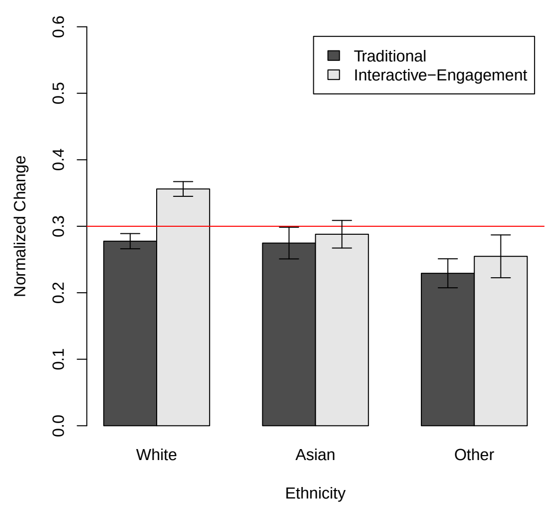
	c_T	c_{IE}
White	0.30 ± 0.01	0.36 ± 0.01
Asian	0.27 ± 0.02	0.31 ± 0.02
Other	0.23 ± 0.03	0.28 ± 0.02
Δ_{W-A}	0.03 ± 0.02	0.05 ± 0.02
Δ_{W-O}	0.07 ± 0.02	0.08 ± 0.02
Δ_{A-O}	0.04 ± 0.03	0.04 ± 0.03

Figure 4.15: CSEM-measured learning gains by ethnicity and modality, IPPS

Plot (a) is the learning gain as measured by the normalized gain. Plot (b) is the learning gain as measured by the normalized change.



(a)



(b)

In the IPLS sequence, learning gains were observed across all ethnicities for both the normalized gain and normalized change (Table 4.13). The gap between White and

Asian students was eliminated with respect to the normalized gain, while a slight increase was noted between White and Other students for both normalized change and normalized change. In Figure 4.16, there was a significant upward trend for all ethnicities and all metrics. No groups, however, exceeded the Hake moderate-gain standard.

Table 4.13: CSEM-measured learning gains by ethnicity and modality, IPLS

Table (a) is the normalized gain analysis. Table (b) is the normalized change analysis.

(a)

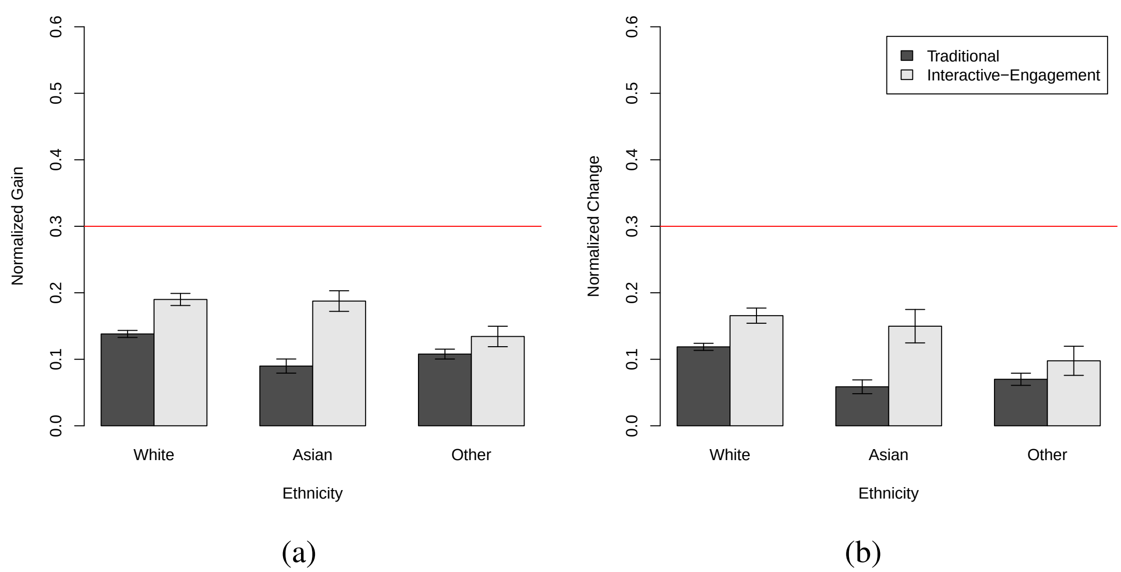
	$\langle g \rangle_T$	$\langle g \rangle_{IE}$
White	0.14 ± 0.01	0.19 ± 0.01
Asian	0.09 ± 0.01	0.19 ± 0.01
Other	0.10 ± 0.01	0.13 ± 0.01
Δ_{W-A}	0.05 ± 0.02	0.00 ± 0.02
Δ_{W-O}	0.04 ± 0.02	0.06 ± 0.04
Δ_{A-O}	-0.01 ± 0.02	0.06 ± 0.04

(b)

	c_T	c_{IE}
White	0.12 ± 0.01	0.17 ± 0.01
Asian	0.06 ± 0.01	0.15 ± 0.02
Other	0.07 ± 0.01	0.10 ± 0.02
Δ_{W-A}	0.06 ± 0.02	0.02 ± 0.03
Δ_{W-O}	0.05 ± 0.02	0.07 ± 0.03
Δ_{A-O}	0.01 ± 0.02	0.05 ± 0.03

Figure 4.16: CSEM-measured learning gains by ethnicity and modality, IPLS

Plot (a) is the learning gain as measured by the normalized gain. Plot (b) is the learning gain as measured by the normalized change.



ANOVAs were conducted across course types (IPLS versus IPPS) and gain metric (normalized gain versus normalized change). In each, the gain metric served as the dependent variable, while the modality and ethnicity factors served as the independent variables. Statistically significant differences in means were found in IPPS courses with respect to the normalized gain only for ethnicity ($F(2,1470) = 7.83$, $p < 0.001$). The main effect of modality was not found to be statistically significant in this model ($F(1,1470) = 2.19$, $p = 0.14$). There was no observed interaction effect between modality and ethnicity. Post hoc testing using the Tukey HSD test found significant differences between White and Other students in the traditional courses ($p < 0.05$), while there was no significant difference in performance between White and Asian students, or Asian and Other students.

Similarly, ANOVA analysis of the IPPS course with respect to the normalized change found both modality ($F(1,1470) = 5.32$, $p < 0.05$) and ethnicity ($F(2,1470) = 8.49$, $p < 0.001$) to have a statistically significant difference in means. There was no observed interaction effect. As with the normalized change, no significant differences between White and Other students in the traditional courses were identified ($p < 0.05$), while there was no significant difference in performance between White and Asian students, or Asian and Other students. This statistically significant result supports the findings in Table 4.12 and Figure 4.15 of an overall increase in learning gains associated with the implementation of interactive-engagement physics education in IPPS courses, while also verifying the increase in the performance gap between White and Asian students and White and Other students.

For the IPLS course, ANOVA analysis of the normalized gain found statistically significant differences in modality ($F(1,3740) = 29.32$, $p < 0.001$) and ethnicity ($F(2,3740) = 8.25$, $p < 0.001$), with no observed interaction effect. Post hoc testing revealed statistically significant differences between the Other and Asian groups ($p < 0.05$), no statistically significant differences were identified between White and Other groups or White and Asian groups. Similar results were observed for the normalized change with modality

($F(1,3740) = 22.09$, $p < 0.001$) and ethnicity ($F(2,3740) = 12.10$, $p < 0.001$). The statistically significant differences in means support the findings in Table 4.13 and Figure 4.16 of an overall increase in normalized gain associated with the implementation of interactive-engagement learning in IPLS courses as well as a decrease in the performance gap between White and Asian and White and Other students in IPLS courses.

5 RESULTS AND CONCLUSIONS

The purpose of this project was to determine the effectiveness of interactive engagement in introductory physics education at the University of North Carolina at Chapel Hill. Effectiveness is determined by studying students' learning gains, and by understanding these learning gains as a function of both gender and ethnicity. To perform these analyses, results from concept inventories administered over seven years were combined with data from the University Registrar to connect test scores with demographic information. This project focused on three core questions, addressed below.

5.1 Has the implementation of interactive-engagement pedagogy at UNC-CH led to student gains in understanding?

Examination of scores from the Force Concept Inventory (FCI) and Conceptual Survey of Electricity and Magnetism (CSEM) from Fall 2008 until Fall 2015 show positive results in both normalized gain and normalized change for all courses (Figure 4.3, Figure 4.4, Figure 4.11, and Figure 4.12). For the introductory physics courses for physical science majors (IPPS courses – Figure 4.3), there was a consistent increase in scores from the beginning of the study period to the present, with a particular jump in Fall 2010, likely corresponding with the introduction of SCALE-UP. Another jump was evident in Fall 2014, the term where lecture/studio and the new content were fully implemented into the curriculum.

FCI results in the introductory physics for life science courses (IPLS – Figure 4.4) also show a consistent increase across the study period, although the reason for this is less clear. Interactive-engagement for these courses only began in Fall 2014, and there was an evident decrease in this semester. It is believed that the consistent gain prior to Fall 2014 is a combination of two effects: the department's focus on improving the introductory curriculum, and an instructor who began implementing basic interactive-engagement concepts in IPLS courses beginning in Fall 2012.

An analysis of variance (ANOVA) analysis considering only course modality on the FCI indicated a very strong statistically significant main effect on modality ($p < 0.01$), indicating that the higher mean scores associated with the interactive-engagement courses are likely to be associated with the change in modality rather than random performance variations. This finding was consistent for courses in both the IPPS and IPLS sequences, providing the strongest evidence that the interactive-engagement pedagogy is effective.

The analysis of scores on the CSEM is more challenging, primarily due to the nature of the CSEM itself. Unlike the FCI, the CSEM is considered to be much more reliant on prior knowledge, so there is a much broader variation in CSEM scores between students in the IPPS sequence versus students in the IPLS sequence. This is usually due to the fact that students in the IPPS sequence have a stronger high school physics background. Figure 4.11 shows the semester-to-semester normalized gain and change in the IPPS sequence – there is no discernible upward trend, but most semesters exceed the Hake moderate-gain threshold normally associated with interactive-engagement courses, especially after the implementation of interactive-engagement in Fall 2010.

The results for the CSEM in the IPLS sequence (Figure 4.12) demonstrate the gap in learning gains compared to the IPPS sequence – the majority of semesters result in a normalized gain $0.05 \leq \langle g \rangle \leq 0.15$. However, a significant increase in normalized gain is observed in Spring of 2015, the semester that interactive-engagement was implemented in the IPLS sequence. Further investigation will need to be conducted to determine if this is a true effect.

Regardless of variability, a statistically significant main effect of modality was identified using an ANOVA analysis for scores on the CSEM for courses in both the IPLS and IPPS sequences ($p < 0.05$). This indicates that the higher average scores associated with the interactive-engagement courses are likely associated with the change in modality. Continued data collection should be able to produce a stronger effect.

5.2 Has the implementation of interactive-engagement pedagogy at UNC-CH led to a significant change in understanding with regard to race or gender?

Results regarding race and gender were unclear. All ANOVA procedures conducted with this research question examined both modality and race or gender. All analyses found a statistically significant effect with regard to modality, and a statistically significant effect with regard to race or gender. Only one analysis found an interaction effect – the modality/gender analysis of the FCI in the IPLS sequence. In other words, for the majority of the analyses there was a clear difference in means between the traditional and interactive-engagement courses, and there was a clear difference in means between male/female or White/Asian/Other – but the analysis was unable to determine whether there were significant differences in means with regard to both factors simultaneously. In essence, an interaction would have better answered the question of whether interactive-engagement pedagogy closes the performance gap between groups. For the one ANOVA where an interaction was noted, the group counts (cell counts) were too unbalanced for further analysis.

However, there were clear improvements across all groups with the implementation of interactive-engagement courses. For the FCI in the IPPS courses, both male and female students exceeded the Hake moderate-gain threshold, while male students exceeded the threshold in the IPLS sequence and female students were within the margin of error in the IPLS sequence. The gender-gap metric Δ_{M-F} narrowed with the transition to interactive engagement for the IPPS sequence, while widening for the IPLS sequence, although the statistical significance of these metrics is unclear. Similar findings were identified in the CSEM for the IPPS and IPLS sequences, with gender gap narrowing in the IPPS sequence and widening for the IPLS sequence. One hypothesis for this effect was the large majority of female students in the IPLS sequence and the possibility of a larger standard deviation, as a broader score distribution could cause this discrepancy. However, the female students had a smaller associated standard deviation with respect to male students, so this theory was discounted. A similar gap narrowing with regard to ethnicity (Δ_{W-A} and Δ_{W-O}) was found in the CSEM results in the IPPS sequence, with the gap widening for the CSEM in the IPLS sequence.

5.3 How do the results at UNC-CH compare to other institutions?

Comparing the results from UNC-CH and other similar institutions is difficult. The General Administration of the University of North Carolina identified fifteen peer institutions for UNC-CH. A listing of these institutions is included as Appendix B of this thesis. All of these peer institutions have colleges or schools of engineering – UNC-CH does not. Although these peer universities are used for institution-wide comparison and not specific to Physics and Astronomy, they illustrate how unusual UNC-CH is in this regard. The lack of an engineering program has a direct effect on the number and quality of enrolled physics students: introductory physics courses are a requirement for engineering programs, and engineering students are more likely to have completed higher levels of physics in high school.

However, comparison of our results to one study (Madsen et al. 2013) is especially encouraging. In this study, concept inventory scores were examined with respect to student gender, and found FCI normalized gains for interactive engagement courses ranging from $\langle g \rangle \approx 0.3$ (U Minnesota, Michigan State) to $\langle g \rangle \approx 0.7$ (Harvard) – UNC-CH's normalized gain for the FCI in the IPPS sequence is approximately $g \approx 0.39$, so our performance is equivalent to other universities, even with our lack of an engineering program. More promising, however, is a comparison of Δ_{M-F} with these institutions – our gap of $\Delta_{M-F} = 0.04$ is less than every school in the survey with the exception of Harvard University and the University of Hull (UK), where female students average score exceeded the average score of the males.

5.4 Future Work and Conclusion

In terms of future research, continued data collection will increase the statistical power of these analyses. It is also highly suggested that these data be analyzed using an Analysis of Covariance (ANCOVA) technique, which could allow for the relative performance with respect to gender and ethnicity to be analyzed in more detail. There would be value in an analysis of test question answers with the FCI and CSEM to determine whether student performance has improved with curriculum and methodology changes. Finally, it could be useful to interview students to collect qualitative data on the efficacy of interactive-engagement in introductory physics courses.

In conclusion, this study finds that the implementation of interactive-engagement introductory physics courses at the University of North Carolina at Chapel Hill has a significant effect on student learning gains as measured by the Force Concept Inventory and the Conceptual Survey of Electricity and Magnetism. This study also found that, for introductory courses for physical science majors, performance gaps narrowed between Male and Female students, White and Asian students, and White and Other students. As interactive-engagement methodologies are continued in introductory courses for life science majors, it is likely that the same gap narrowing results will be identified. There is a clear positive effect on learning gains in introductory physics courses after the implementation of interactive-engagement pedagogies, demonstrating that UNC-CH's efforts to improve introductory physics courses were successful. Furthermore, the results of this study suggest that the implementation of interactive-engagement pedagogies would be beneficial to other institutions seeking to improve teaching and learning in introductory physics.

A INTRODUCTORY PHYSICS COURSE TOPICS

This appendix includes topic lists from all introductory physics courses in the IPLS and IPPS sequences. This list should not be all inclusive, as it is based off of instructor schedules and syllabi.

Table A.1: Course topics: IPLS Mechanics (PHYS 104/114)

<table>
<tr><th>PHYS 104</th><th>PHYS 114</th></tr>
<tr><td>Measurement
Kinematics
Vectors
Relative motion
Newtons Laws
Weight and normal forces
Work and kinetic energy
Potential energy
Momentum and collisions
Rotational kinematics
Rotational dynamics
Gravity and potential energy
Oscillations about equilibrium
Waves and Sound
Fluids
Temperature and Heat
Phases and Phase Changes
Thermodynamics</td><td>Estimations and scaling
Kinematics
Newtons Laws
Stress and Strain
Torque
Force, Work and Kinetic Energy
Gravitational Potential Energy
Energy Conversion and Elastic Potential Energy
Oscillations
Sound
Ideal Gases and Heat Capacity
Diffusion
Heat Transfer</td></tr>
</table>

Table A.2: Course topics: IPPS Mechanics (PHYS 116/118)

PHYS 116	PHYS 118
• Measurement • Velocity and acceleration • Accelerated motion • Vectors • Relative motion • Projectile Motion • Circular Motion • Newtons Laws • Energy and Work • Potential Energy • Energy conservation • Collisions • Rotational kinematics • Potential energy and Keplers Laws • Fluids • Oscillations, SHM • Wave motion • Sound, Doppler effect • Relativity (in Lecture/Studio 116)	• Motion • Kinematics • Vectors • Relative motion • Rotational kinematics • Forces and motions • Dynamics • Friction • Newtons Laws • Impulse and momentum • Kinetic and Potential Energy • Hookes Law • Work, conservation of energy • Relativity (Lorentz transformations, proper time and length) • Relativistic momentum and energy • Rotational dynamics • Gravity and satellites • Oscillations and SHM • Waves • Sound, light, Doppler effect

Table A.3: Course topics: IPLS Electricity and Magnetism (PHYS 105/115)

PHYS 105	PHYS 115
• Electric charges • Electric forces and fields • Electric potential • Electric current and circuits • Magnetic forces and fields • Electromagnetic waves • Geometrical optics • Physical optics • Relativity • Quantum physics • Atomic physics • Nuclear physics	• Fluids • Electric charges • Electric forces and fields • Electric potential • Electric current and circuits • Magnetic forces and fields • Geometrical optics • Wave optics • Thin films • Nuclear physics

Table A.4: Course topics: IPPS Electricity and Magnetism (PHYS 116/118)

PHYS 116	PHYS 118
• Electric charges • Electric forces and fields • Gauss Law • Electric potential • Capacitance • Resistance • Electric circuits • Magnetic fields and Lorentz forces • Biot-Savart law and Amperes Law • Lenz Law • Inductors and transformers • Maxwells Equations • Magnetic materials • Geometrical optics	• Charges • Electrostatics • Electric forces and fields • Gauss Law • Electric potential • Capacitance • Resistance and Current • Electric Circuits • Magnetic fields and Biot-Savart Law • Lorentz forces and LenzLaw • Induction and AC circuits • Maxwells Equations • Geometrical optics • Interference • Photoelectric effect and Photons • Matter-Waves and quantization • Probability and Wavefunctions • Uncertainty principle • 1-D quantum mechanics • Atomic physics • Nuclear physics

B UNC-CHAPEL HILL PEER INSTITUTIONS

The General Administration of the University of North Carolina system defines the following institutions as “peer institutions” for the purposes of institutional research and assessment:

- University of California - Berkeley
- University of California - Los Angeles
- University of Maryland - College Park
- University of Michigan - Ann Arbor
- University of Minnesota - Twin Cities
- University of Pittsburgh
- University of Texas at Austin
- University of Virginia
- University of Washington - Seattle
- University of Wisconsin - Madison
- Duke University
- Johns Hopkins University
- Northwestern University
- University of Pennsylvania
- University of Southern California

REFERENCES

Arons, A. B. (1981). Thinking, reasoning, and understanding in introductory physics course. *The Physics Teacher*, 19:166–172.

Bao, L. (2006). Theoretical comparisons of average normalized gain calculations. *American Journal of Physics*, 74:917.

Beichner, R. J. (1999). Introduction to the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) project. In *Reshaping Undergraduate Science and Engineering Education: Tools for Better Learning*, Sigma Xi Forum, pages 44–52. Sigma Xi.

Beichner, R. J. and Saul, J. M. (2004). Introduction to the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) project. In Vicentini, M. and Redish, E., editors, *Proceedings of the International School of Physics "Enrico Fermi"*.

Churukian, A. D. (2002). *Interactive Engagement in an Introductory University Physics Course: Learning Gains and Perceptions*. PhD thesis, Kansas State University.

Churukian, A. D. (2016a). Personal correspondence, April 2016 e-mail.

Churukian, A. D. (2016b). Personal correspondence, February 2016 e-mail.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50:66.

Clement, J., Brown, D. E., and Zietsman, A. (1989). Not all preconceptions are misconceptions: finding 'anchoring conceptions' for grounding instruction on students' intuitions. *International Journal of Science Education*, 11:554–565.

Dewey, J. (1938). *Experience and Education*. Simon and Schuster.

Driver, R. (1995). Constructivist approaches to science teaching. In Steffe, L. and Gale, J., editors, *Constructivism in Education*, pages 398–399. Lawrence Erlbaum, Hillsdale, NJ.

Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. Sage Publications, Los Angeles, CA.

Fischbein, H. (1987). *Intuition in science and mathematics: an educational approach*.

Springer.

Fisher, K. M. (2004). The importance of prior knowledge in college science instruction. In Sunal, D. W., Wright, E. L., and Day, J. B., editors, *Reform in Undergraduate Science Teaching for the 21st Century*. Information Age Publishing.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition.

Furth, H. (1969). *Piaget and Knowledge*. Prentice-Hall, New York.

Gergen, K. J. (1995). Social construction and the educational process. In Steffe, L. and Gale, J., editors, *Constructivism in Education*, pages 398–399. Lawrence Erlbaum, Hillsdale, NJ.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66:64.

Halloun, I. A. and Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, 53:1043.

Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30:141–158.

Jardine, D. W. (2006). *Piaget & Education Primer*. Peter Lang, New York.

Johnson, D. W., Johnson, R. I., and Smith, K. A. (1997). Academic controversy. enriching college instruction through intellectual conflict. Technical report, Association for the Study of Higher Education.

Karplus, R. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching*, 14(2):169–175.

Knight, R. D. (2004). *Five easy lessons : strategies for successful physics teaching*. Addison Wesley, San Francisco, Calif. ID: 54484133.

Lawson, A. E. (1982). The nature of advanced reasoning and science instruction. *Journal of Research in Science Teaching*, 19(9):743.

Lawson, A. E. (2005). What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching*, 42(6):716–740.

Madsen, A., McKagan, S. B., and Sayre, E. C. (2013). Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics - Physics Education Research*, 9(020121).

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., and Heuvelen, A. V. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, 69:S12.

Mann, C. R. (1912). *The Teaching of Physics for the Purposes of General Education*. Macmillan, New York.

Marx, J. D. and Cummings, K. (2007). Normalized change. *American Journal of Physics*, 75(1):87–91.

Marzano, R. J. (2007). *Art and Science of Teaching: A Comprehensive Framework for Effective Instruction*. Association for Supervision and Curriculum Development, Alexandria, VA.

McCormick, B. (2004). Science education reform and higher education: A historical perspective. In Sunal, D. W., Wright, E. L., and Day, J. B., editors, *Reform in Undergraduate Science Teaching for the 21st Century*, page 17. Information Age Publishing, Greenwich, CT, 1st edition.

McDermott, L. C. and Redish, E. F. (1999). Resource letter: PER-1: Physics Education Research. *American Journal of Physics*, 67(9):755–767.

McKinnon, J. W. and Renner, J. W. (1971). Are colleges concerned with intellectual development. *American Journal of Physics*, 39:1047.

McNeil, L. E. (2004). Transforming introductory physics teaching at UNC-CH. http://user.physics.unc.edu/ mcneil/physicsmanifesto.html.

Meltzer, D. E. and Otero, V. K. (2015). A brief history of physics education in the united states. *American Journal of Physics*, 83:447.

Moore, J. C. (2012). Transitional to formal operational: Using authentic research experiences to get non-science students to think more like scientists. *European Journal of Physics Education*, 3(4).

Otero, V. K. (2004). Cognitive processes and the learning of physics part i: The evolution of knowledge from a vygotskian perspective. In Vicentini, M. and Redish, E., editors, *Proceedings of the International School of Physics "Enrico Fermi"*.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Redish, E. F. (2003). *Teaching Physics with the Physics Suite*. Wiley, New York.

Renner, J. W. and Lawson, A. E. (1973). Piagetian theory and instruction in physics. *The Physics Teacher*, 11:165.

Seymour, E. (1995). Revisiting the 'problem iceberg': science, mathematics, and engineering students still chilled out. *Journal of College Science Traching*, 24(6):392. doi: pmid:.

Sorensen, C., Churukian, A., Maleki, S., and Zollman, D. (2006). The new studio format for instruction of introductory physics. *American Journal of Physics*, 74:1077.

Tabachnick, B. and Fiddell, L. (2007). *Experimental Designs Using ANOVA*. Duxbury, Belmont, CA.

The ATLAS Collaboration (2012). Observation of a new particle in the search for the standard model higgs boson with the {ATLAS} detector at the {LHC}. *Physics Letters B*, 716(1):1 – 29.

von Glasersfeld, E. (1995). A constructivist approach to teaching. In Steffe, L. and Gale, J., editors, *Constructivism in Education*. Lawrence Erlbaum, Hillsdale, NJ.

Weisman, L. I. and Safford, P. L. (1971). Piagetian tasks as classroom evaluative tools. *The Elementary School Journal*, 71(6):329–338.

Zollman, D. (2004). A model for reform in teaching physics: Large-enrollment physics classes. In Sunal, D. W., Wright, E. L., and Day, J. B., editors, *Reform in Undergraduate Science Teaching for the 21st Century*, page 17. Information Age Publishing, Greenwich, CT, 1st edition.