AN ABSTRACT FOR A DISSERTATION

CONCEPTUAL PHYSICS DIFFERENCES BY PEDAGOGY AND GENDER: QUESTIONING THE DEFICIT MODEL

Twanelle Deann Walker Majors

Doctor of Philosophy in Exceptional Learning STEM

The differences in physics performance between males and females have been studied extensively (Blue & Heller, 2003; Coletta, 2015; Madsen, McKagan, & Sayre 2013; McCullough, 2002, 2004, 2011; Pollock, Finkelstein, & Kost, 2007; Zohar & Sela, 2003). The purpose of this study was to look at the ways teaching methods and assessment choices have fabricated a gender gap. Deficit ways of thinking have further marginalized women by renegotiating prior acts of power that initiated and perpetuated marginalization. Outside of the deficit model, the blame for the underperformance of females has been attributed to discourses of power as well as less-than-critical ways of evaluating learning and schooling. Students in introductory algebra-based physics courses from 2008–2014 at Tennessee Technological University were self-enrolled in PHYS2010 sections that were taught using either a traditional or constructivist, interactive-engagement Learner-centered Environment for Algebra-based Physics (LEAP) pedagogy. Propensity scoring on all feasible and relevant independent variables was used to adjust for the probability of students choosing either LEAP or traditional sections. The Force Concept Inventory (FCI) and Gender Force Concept Inventory (GFCI) were used as the measures to gauge students' performance on physics concepts. The results showed that there were no differences in the FCI or GFCI performance of males and females. Results also showed that when accounting for pretest performance and the likelihood of choosing a LEAP section, LEAP pedagogy accounted for roughly 30% of performance differences. Not only was this true on the average, it was true for both genders. This meant that the main effect of LEAP pedagogy was even stronger and more generalizable. Gender did not moderate pedagogy, indicating that a pedagogy gap focus was more appropriate for evaluating physics learners.

CONCEPTUAL PHYSICS DIFFERENCES BY PEDAGOGY AND GENDER: QUESTIONING THE DEFICIT MODEL

A Dissertation

Presented to

the Faculty of the College of Graduate Studies

Tennessee Technological University

by

Twanelle Deann Walker Majors

In Partial Fulfillment

of the Requirements of the Degree

DOCTOR OF PHILOSOPHY

Exceptional Learning STEM

May 2015

CERTIFICATE OF APPROVAL OF DISSERTATION

CONCEPTUAL PHYSICS DIFFERENCES BY PEDAGOGY AND GENDER: QUESTIONING THE DEFICIT MODEL

by

Twanelle Deann Walker Majors

Graduate Advisory Committee:

Holly Anthony — 4-20-15
Holly Garrett Anthony, Co-chairperson — Date

George Chiuyo — 4/20/15
George Chiuyo, Co-chairperson — Date

David Larimore — 4/20/15
David Larimore — Date

Martha J. Howard — 4-20-15
Martha Howard — Date

Paula V Engelhardt — 4-20-15
Paula Engelhardt — Date

S T Robinson — 4/20/15
Stephen Robinson — Date

Approved for the Faculty:

Mark Stephens
Mark Stephens
Dean
College of Graduate Studies

5/5/15
Date

DEDICATION

This is for the girls, females, ladies, and women. This is mostly for Granny Walker (show no mercy) and Granny Lucille (look both ways), both great people.

I realize that my mother and grandmothers put their lives aside on several occasions to see that I could go on. I dedicate this to you all for showing my children that "going on" is a good thing for a mother to do. To Treeah, thank you for being the person I could count on for stimulating discussions on social justice matters in academia. This work has often reminded me of my own college physics experience, looking forward to you waking in the night and being my tiny study buddy or study break. I am sure that is when imprinting of neurotic study habits was concretized. To Chloe, I cannot measure the impact you have had on my choice to pursue this degree and my feelings that I would see it to the end. Without your company during the late study hours, this would have been a chore rather than a chance to watch you grow to be a chemistry nerd. To Rily, you are my favorite man in this world. Thank you for getting fighting mad on my behalf when there were barriers in this process and for insightful observations about the biases on assessments. To my dissertation coach, Ginger, thank you for being my research assistant and for motivating me with your solid study habits in the office/closet. Your discussions on the unjust schooling practices endured by students with special needs or language differences inspired many moments in this process.

ACKNOWLEDGEMENTS

In 1999, Dr. Margaret Phelps suggested I join a physical science institute during which Dr. Steve Robinson modeled a constructivist, interactive-engagement pedagogy. Thank you both for changing how I thought science should be taught. To Drs. Paula Engelhardt and Robinson, thanks for letting me join your amazing project and for tolerating my love of rigorous confusion. Dr. Engelhardt, thank you for sharing your experiences as a woman in STEM and for recognizing when I was overcome by it. I have truly enjoyed working with you. Dr. Holly Anthony, thank you for combing this work so thoroughly, introducing me to alternative representations, and being genuinely concerned that the end product told the story well. I thank Dr. Lisa Zagumny for substantial feedback and for being the first female (non-relative) to encourage my critical perspective. Dr. Martha Howard, thank you for making me want to question objectivity and for my first supportive cohort experience. I will forever (re)shift, and feel safe to do so, because of it. Dr. David Larimore, thank you for bringing candor to this work and being the voice of striking rhetoric. Dr. George Chitiyo, you are the most reliable and responsive mentor ever, and my work with others will forever reflect your example. Also, thank you for giving your accent to the statistical voice in my mind.

I thank Moma for countless hours of help with data entry. To Daddy and Lang, thanks for always assuming I was nearly done and for being men who were good to my children. I especially thank Moma and Granny Walker for helping rear my children. Without that form of support, this would not have happened. Most of all, I thank my four children for being a supportive and inspiring group of fine people.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF ABBREVIATIONS ix

CHAPTER 1 - INTRODUCTION 1

- Statement of the Problem 1
- Statement of Purpose & Research Questions 2
- Significance of the Study 4
- Delimitations 7
- Definition of Terms 8
  - Diagnostic Test 8
  - Deficit Ideology 9
  - Gender Biased Assessment 9
  - Hake Gain 9
  - Interactive Engagement 10
  - LEAP 10
  - Normalized Change 11
  - Normalized Gain 11
  - Powerblind 11
  - Reformed Teaching 12
  - Traditional Physics Course 12
- Theoretical Perspective 13
  - Positivism 13
  - Paradigm (Re)Shift 14
  - Postmodernism 15
  - Critical Theory 18
- Subjectivities: Influential Coursework 19
- Research Assumptions 21
- Summary 21

CHAPTER 2 - REVIEW OF THE LITERATURE 24
Deficit Ideology 24
The Deficit Model of Achievement 25
Synthesis of Deficit/Gap Literature 28
Females, Girls, Ladies, Women & Schooling 29
History of STEM Education 34
Gender & STEM Education 36
Gender & Mathematics Education 38
Gender & Science Education 39
Gender & Physics Education 41
Pedagogy & Pedagogical Knowledge 43
Traditional Pedagogies 46
Reformed Teaching Pedagogy 48
Force Concept Inventory 50
Force Concept Inventory Development 50
Field Testing of the FCI 51
Significance of Prior FCI Studies 53
Deficits in Instruction 53
Context of Culture 54
Gender Force Concept Inventory 55
Gender Force Concept Inventory Development 56
Field Testing of the GFCI 57
Summary 58
CHAPTER 3 - METHODOLOGY 59
Introduction 59
Research Design 60
Context of the Study 60
Population & Sample 62
Instruments 64
Force Concept Inventory Validity & Reliability 64
Gender Force Concept Inventory Validity & Reliability 70

Methods 71
Pre-Analysis Data Screening 71
Data Analysis 73
Summary of Methods 75
CHAPTER 4 - DATA ANALYSIS 76
Pre-Analysis Data Screening 76
Variable Descriptions 76
Missing Cases 79
Propensity Scoring 79
Results 84
Force Concept Inventory Analysis 84
Gender Force Concept Inventory Analysis 88
Concept Inventory Factor Analysis 92
Testing for Differences on Force Concept Inventory Constructs 92
Testing for Differences on Gender Force Concept Inventory Constructs 98
CHAPTER 5 - INTERPRETATION 99
THE PEDAGOGY GAP! SOMEONE OWES WOMEN AN APOLOGY! 99
Significance of Findings 100
Pedagogy 100
Gender 101
Pedagogy & Gender 102
Recommendations 103
Pedagogy 103
Gender 104
Synthesis of the Findings 105
Limitations 106
Discussion 107
REFERENCES 111
BIBLIOGRAPHY 123
APPENDIX – CODING OF MAJORS & PROGRAMS 126
VITA 128

LIST OF TABLES

Table	Page
1. Summary of Analysis Methods Used to Answer Research Questions	74
2. Logistic Regression Coefficients	82
3. Logistic Regression Classification Table with No Predictors	83
4. Logistic Regression Classification Table with Predictors	83
5. Unadjusted FCI Means by Pedagogy	84
6. Unadjusted FCI Means by Gender	84
7. ANOVA Summary for FCI Normalized Gain Blocked on Propensity Score	86
8. ANOVA Summary for FCI Posttest Blocked on Propensity Score & Pretest	88
9. Unadjusted GFCI Means by Pedagogy	89
10. Unadjusted GFCI Means by Gender	89
11. ANOVA Summary for GFCI Normalized Gain Blocked on Propensity Score	90
12. ANOVA Summary for GFCI Posttest Blocked on Propensity Score & Pretest	92
13. Force Concept Inventory Factor Loadings	95
14. Pedagogy Differences on FCI Factors	96
15. Gender Differences on FCI Factors	97

LIST OF ABBREVIATIONS

Abbreviation	Description
AAAS	American Association for the Advancement of Science
ACT	American College Testing
ANOVA	Analysis of Variance
AP	Advanced Placement
DV	Dependent Variable
FCI	Force Concept Inventory
FMCA	Force and Motion Conceptual Evaluation
GFCI	Gender Force Concept Inventory
GPA	Grade Point Average
IV	Independent Variable
LEAP	Learner-centered Environment for Algebra-based Physics
MB	Mechanics Baseline
MD	Mechanics Diagnostic
NCES	National Center for Educational Statistics
PER	Physics Education Research
PET	Physics and Everyday Thinking
STEM	Science, Technology, Engineering, and Mathematics
TTU	Tennessee Technological University

CHAPTER 1

INTRODUCTION

Statement of the Problem

If teaching were found to be unrelated to learning, for any group, education would arguably have been the largest-scale compulsory form of hazing. For education policy to promote social justice for all, data and research that informed policy must have been contextually grounded. Education policy has often been driven by national priorities and concerns that may have been informed by an incomplete or inaccurate assessment of status and needs. Global competitiveness in Science, Technology, Engineering, and Mathematics (STEM) has continued to be a national priority. Student performance in STEM was a critical concern in recent education reform movements. Past studies have shown differences in STEM performance to be related to gender, with females performing lower on measures of achievement (Else-Quest, Hyde, & Linn, 2010; Madsen et al., 2013; Pollock et al., 2007). Researchers have explored background differences of males and females as possible reasons for lower performance of females in physics (Kost, Pollock, & Finkelstein, 2009; Kost-Smith, 2011; McCullough, 2002). Nontraditional teaching methods have been shown to narrow the gender gap in physics (National Research Council, 2012). These studies often offered no transparency to the positivist framework that informed and constrained the work, a behavior that has been endorsed by organizations which purport to represent research as a discipline (Harding, 1993b; Lather, 2004b). This powerblind ideology in science education research has propagated and

perpetuated empirical research—and thus common practice—that lacks strong objectivity (Harding, 1993a). Past studies have pointed to gender bias of the Force Concept Inventory (FCI), a limitation in interpretability of teaching methodology research, though it continued to be the most popular measure of physics conceptual knowledge (Hestenes, Wells, & Swackhamer, 1992). What is not fully understood is whether the differences between male and female performance in physics were mediated by teaching method and moderated by gender. Said another way, the underperformance of females may have been related to teaching method and assessment choice rather than deficits associated with being a female in a physics classroom.

If gender differences can be attributed to the context of learning and assessment, deficit models of thinking can be further put to rest. For any multicultural classroom, deficit thinking only blamed females for lacking characteristics of maleness being measured. Under a deficit model of thinking about the lackluster performance of females in physics, several characteristics have been proposed as reasons for differences: mathematics ability, previous coursework, and attitudes (Kost et al., 2009; Lorenzo, Crouch, & Mazur, 2006; McCullough, 2002). Some studies have reported contradictory results suggesting that there are no differences between male and female performance on the FCI (Blue & Heller, 2003; Kost-Smith, 2011).

Statement of Purpose & Research Questions

The purpose of this study was to determine the differences between students who self-enrolled in either a traditional algebra-based introductory physics or nontraditional

algebra-based introductory physics in performance on the FCI and a female-centric version of the FCI—the Gender Force Concept Inventory (GFCI). In addition, this study determined whether there was a significant difference between male and female students of traditional and nontraditional algebra-based physics in performance on the FCI and GFCI. Finally, this study determined if any differences associated with pedagogy (traditional versus Learner-centered Environment for Algebra-based Physics (LEAP)) were consistent for males and females.

This study addressed the following questions. All questions addressing differences (questions 1–6) included a control variable of predicted group membership based on the propensity scores:

1. Is there a difference in the performance of males and females on the FCI?
2. Is there a difference in the performance of males and females on the GFCI?
3. Is there a difference between students taught using LEAP pedagogy versus a traditional pedagogy on the FCI?
4. Is the difference in performance on the FCI between students who were in the LEAP vs. traditional the same for males and females?
5. Is there a difference between students taught using LEAP pedagogy versus a traditional pedagogy on the GFCI?
6. Is the difference in performance on the GFCI between students who were in the LEAP vs. traditional the same for males and females?
7. Is there a difference between males and females on particular constructs of the FCI?

8. Is there a difference between students taught using LEAP pedagogy versus a traditional pedagogy on particular constructs of the FCI?

Significance of the Study

Establishing a robust link in the chain, rather than attempting to determine all links at once, to get buy-in from teachers, administrators, and policy makers is key to a transformed multicultural STEM education. Because white males have been reported to outperform other groups in STEM, two explanations came to mind: 1) other groups lacked characteristics of white maleness necessary for performance in STEM or 2) STEM was currently oriented for white males. This study investigated practices that were possible robust links in the chain of events that were related to STEM performance for all groups. Rather than attribute performance to group characteristics, this study explored teaching methods and assessment bias as practices that possibly created artificial performance differences.

The FCI has been a gold standard in the Physics Education Research (PER) community for two decades, though intra-journal debates have continued to be numerous and varied as to which characteristic of the instrument was dissected (Hake, 2007; Heller & Huffman, 1995; Hestenes & Halloun, 1995; Wang & Bao, 2010). The GFCI was developed by Laura McCullough as a female-centric version parallel to the male-centric FCI (McCullough & Foster, 2001). Few researchers have reported the use of the GFCI as an alternative for ongoing physics education research. This study added to the PER community by initiating a conversation about the use of instruments which have

substantiated gender bias and the need for caution in citing relationships between teaching methods, gender, and physics performance that were measured using a biased assessment. This work intended to improve classroom practices, teacher education practices, and professional development practices by describing the relationship between teaching method and performance in science. The evaluation of pre-service teachers was an important venue for conveying the characteristics of reformed teaching practices, as every PHYS2010 student studied here received his or her prior STEM education from a teacher who was a product of a pre-service teacher education program. The work described here was significant in that it described the state of physics understanding that followed K–12 STEM education and offered a disaggregation of the concepts most related to reformed teaching practices. National and local education reform movements have been driven by data in its many forms. These reform movements have informed policy on evaluation, program development, testing, and interventions for underperforming groups. For policy to promote social justice, data and research that informed policy must have been contextual grounded. The methods by which student performance was evaluated in this study modeled a social justice approach to quantitative analysis of student performance.

This study added to test discourse by describing the level of physics conceptual understanding by undergraduates and the relationship between teaching and understanding. This was of national importance given that schooling was one of few compulsory life processes. If teaching was not related to learning—for any group, it was unjust to continue to subject members of the group to schooling. Conceptual test development in chemistry has elicited discourse in the chemistry education research

community regarding the stark differences between what teachers thought their students knew and how little conceptual knowledge was demonstrated from students at a variety of ability levels. Much like the phenomena experienced by FCI researchers, they found that poor recall of a person's past alternate conceptions caused disbelief on part of some faculty, and was rooted in the idea that the test was trivial (Mulford & Robinson, 2002). The test appeared simplistic and easy, but results illuminated deep-seated misconceptions even after substantial amounts of instruction.

Hake (1998) solicited FCI data from high school and college teachers who used the FCI in their own action research, with teachers of 62 courses and over 6,000 students included in the data analysis. None of the courses in his analysis, despite pedagogical approach, showed high Hake gains ($\langle g \rangle \geq 0.7$). This information was used at Tennessee Technological University (TTU) to inform LEAP development. Knowing areas on the FCI for which LEAP students did not outperform traditional students stimulated a look at where the idea was addressed in the curriculum and how to question what teachers were doing. Though small adjustments to curricular materials and course structure at TTU have resulted from data analysis each semester, a significant change in LEAP curriculum took place in fall of 2010. The proposed research informed the further improvement of the LEAP program, as well as offered alternate explanations for the perceived underperformance of females in physics.

This study added to the research on STEM performance by describing how to analyze the effect of an instructional method using a statistical treatment that utilized propensity scoring on a large number of relevant and feasible covariates. Statistical treatment of instructional data has largely resulted in describing relationships between

student characteristics, teaching methods, and performance. Gender differences in STEM performance have been studied using a deficit model in which male characteristics were attributed to the higher performance of males. This work was a rebuttal to the deficit model in that it used background variables to categorize students on the likelihood of choosing either the LEAP or traditional methods course. Adjusting for the propensity to choose a particular teaching method to determine the effect of teaching method on FCI performance has not been reported. This novel method for analyzing FCI performance strengthened the findings in comparison to omitting pre-analysis propensity scoring.

Delimitations

The results of this study could be generalized to students who (a) were enrolled in the first semester of an introductory algebra-based physics course for non-majors, (b) at a public university in Tennessee and (c) had a choice between a traditional or LEAP pedagogy.

The FCI was used as a measure of physics performance because it was established by the community as the gold standard despite studies reporting a gender bias on the FCI. The GFCI was implemented in the Spring 2014 semester in place of the FCI for the purpose of continuing the ongoing departmental pedagogical research. Other measures common to the PER community were not used in this work since they were not consistently used from 2008–2014. The findings of this study were not intended to be generalized to populations being measured by algorithmic, rather than conceptual, physics assessments.

Gender and pedagogy were chosen as the focus for this study of difference in physics performance. Previous math and physics coursework from high school transcripts and college records up to the time of PHYS2010 (an introductory algebra-based physics course at TTU) enrollment, grades in the relevant previous coursework, ACT composite, ACT Mathematics score, ACT science reasoning score, high school GPA, high school location, location of permanent residence, classification at the time of PHYS2010 enrollment, declared major at the time of PHYS2010 enrollment, race, ethnicity, and gender were the appropriate, available covariates from a larger list of covariates. Due to high correlations between some of these variables, the following were used to determine if choosing between a traditional course and a LEAP course was related to background: ACT Mathematics score, ACT science reasoning score, high school GPA, high school location, location of permanent residence, classification at the time of PHYS2010 enrollment, declared major at the time of PHYS2010 enrollment, race, ethnicity, and gender. These covariates were used to create a single variable, the propensity score.

Definition of Terms

Diagnostic Test

Rather than an achievement measure, a diagnostic test provided a description of the type of thought being used to explain a phenomenon (Hestenes, Wells, & Swackhamer, 1992).

Deficit Ideology

Deficit thinking was a method for explaining differences and underperformance by implicating a lack of particular characteristics or qualities as the cause of differences, thus blaming the lackluster performance of females on females as opposed to implicating policies, procedures, and hegemonic practices for creating those differences (Gorski, 2011).

Gender Biased Assessment

An assessment instrument written in a manner that produced differences between males and females that were not attributed to the differences in ability on the criterion intended to be measured was gender biased (McCullough, 2004). Said another way, and particularly germane to this study, the assessment lacked face validity if it did not appear (on its face) to measure what it purported to measure—understanding of the force concept—unless you were male (Kachigan, 1991; Mertler & Vannatta, 2005; Witte & Witte, 2007).

Hake Gain

Performance for an entire course was measured as a ratio of the average student gain to the average maximum possible gain (Marx & Cummings, 2007).

Interactive Engagement

As a signature pedagogy for STEM instruction, interactive engagement was identified by a student-centered course in which minds-on and sometimes hands-on work was done throughout the lesson for the purpose of yielding immediate feedback from the instructor (Hake, 1998).

LEAP

Learner-centered Environment for Algebra-based Physics (LEAP) was a nontraditional, guided inquiry-based two-semester introductory algebra-based physics curriculum suitable for medium enrollment classes with a diverse student population that was relatively under-prepared in terms of their math and scientific reasoning skills. Since 2008 the Physics Department of Tennessee Technological University (TTU) offered at least one LEAP section of PHYS2010 per semester. The LEAP curriculum used the same pedagogical structure as Physics and Everyday Thinking (PET). Since PET was a purely conceptual course (Goldberg, Otero, & Robinson, 2010), LEAP was designed by extending PET structure to incorporate algebraic representations (formulas) and a deliberate problem solving strategy to reinforce links between different representations (S. Robinson & P. Engelhardt, personal communication, April 17, 2015). Ongoing action research indicated that the guided inquiry-based learning in a learner-centered environment was more conducive to the development of a deeper understanding of the

conceptual ideas of force, motion, and energy as applied to mechanics than in traditional lecture with separate exposure-verification laboratories.

Normalized Change

Individual student performance was measured as of a ratio of actual gain to the maximum possible gain in instances where gain was positive, and was measured as a ratio of actual gain to the maximum possible loss where gain was negative (Marx & Cummings, 2007). For this study, the calculation as described by Marx and Cummings was used as a measure of performance. As is common in the PER community, this calculation was called normalized gain rather than normalized change.

Normalized Gain

Individual student performance was measured as of a ratio of actual gain to the maximum possible gain (Marx & Cummings, 2007).

Powerblind

Knowing that decisions related to policy, curriculum, assessment, evaluation, intervention, or access to services were influenced by factors other than the best interests of the population being studied, while carrying on as if no power structures impacted

decisions, was defined here as a powerblind discourse (Kurzman et al., 2014) Germane to this study was Patti Lather's (1991, 2012) argument that those who did science often used a powerblind ideology to disengage from the necessity of taking a hard look at the social constructs and political discourses that influenced the work they are doing. Using a remediation model (arguably a deficit model) to secure funding, rather than a proposal that increased access or opportunities before remediation seemed necessary, was an example of powerblind discourse.

Reformed Teaching

Teaching behaviors that demonstrated constructivist pedagogies within a student-centered interactive engagement classroom environment were considered reformed (MacIsaac & Falconer, 2002).

Traditional Physics Course

For the purpose of this study, a traditional physics course was described as a teacher-centered course, in which students have a passive role, characterized by lectures, recipe laboratory activities, and algorithmic exams (Hake, 1998).

Theoretical Perspective

Positivism

Though the positivist paradigm was the theoretical framework of the analytic methods of this study, a subjective and suspicious lens informed the design of the study and the transparent attempt at strong objectivity (Harding, 1993a). Science and science education have had a long history of knowledge built and maintained by the idea that the process of doing science, through some version of the scientific method, was the best means for getting unproblematic dehumanized information (Bazzul, 2013; Bower, 1998; Lather, 2004b; St. Pierre, 2002). The power discourses engaged in and within science education were hardly recognizable by the populous because of the simplistic and narrow description of what constituted science (Lather, 2012). The objectivity born out of the 19th century was a framework for knowing that was shaped by the success and growth of science, outside pressure to solve societal problems, and the means for justifying research funding (Bower, 1998; Lather, 2004b, 2005; St. Pierre, 2002). Bower (1998) contrasted the viewpoint of the scientists, one which chipped away at reality through perseverance in order to develop theories grounded in real-world evidence and thus a culture-free way of knowing. Since all of our "knowing" occurred within the confines of a culture, this seemed hard to attain. The absence of contextual and cultural critiques in science education pedagogical spaces, and science education research, has been hegemonic and has perpetuated inequities (Bazzul, 2013). Lather (1991, 2012) also argued that those who did science often used a powerblind ideology to disengage from the necessity of taking a

hard look at the social constructs and political discourses that were influencing the work they were doing.

Modern meanings of scientific "objectivity" have included empirical reliability, procedural correctness, emotional detachment, and absolute truth (Bower, 1998). Lyotard (1984) explained the game, by which science allowed and disallowed information to count as knowledge, as a product of the idea that the box was the only way we could truly know—through consensus, objectivity, and narrow but universally applicable methods. Lather (2012) further resonated the work of Lyotard with a call for critical approaches to education research in which subjectivity was valued. She also purported that these attempts to collaborate subjectivity with scientific research "will be initiated primarily by women as men have more to lose" (Lather, 2004b, p. 766).

Paradigm (Re)Shift

Regrettably, the positivist paradigm I was indoctrinated into during my K–12 experience as a student and my college experience as a Biology: Pre-Medicine major had to be questioned and brought to terms in this quantitative study. I was particularly engaged by the work of Patti Lather because of her willingness to focus on things often swept under the rug by a sort of powerblind discourse justified by positivist ideologies. Her critique of the unwillingness to trouble the "hard stories of racism and inequality" (p. 1022) in both multiculturalism and science education movements was a "line of flight" Deleuze & Guittari, 1987, p. 3) that I experienced each time I thought of the research questions necessary to more completely study the barriers in STEM education and STEM

work fields (Lather, 2012). I, too, believed that asking the questions that were often dealt with by avoiding dealing with them (through a positivist design of research) "[got] us nowhere and offer[ed] no useful critique of the shaping influences of the workings of power over education, science or otherwise" (Lather, 2012, p. 1022).

The work of Sipe and Constable (1996, p.156) described four research paradigms in a transgressive but illustrative way. During my first year as a doctoral student, I struggled with understanding my position amongst the paradigms. This piece of literature was important in concretizing my research identity and helped me accept my tendency to use *heat* in discourse. The authors described each paradigm in terms that were research jargon-free. For critical theory, "if this research paradigm were a color, it would be: red (dynamic, action-oriented)" (p. 156) which described the methodologies of this study. The research questions for this study were critical questions about learners and gender, so that "if this research paradigm were a public event, it would be: a March of Dimes telethon (active, purposeful, concerned with marginal groups)" (p. 156). Critical theory was also described in terms of a personality disorder, "it would be: manic-depressive (rage against unjust power structures; bleak worldview with outbreaks of enthusiastic activism)," (p. 156) which did a fairly good job of describing this research process.

Postmodernism

Postmodernism holds that there is not one truth for all, so it has often been thought to be the opposite of objectivism. According to St. Pierre (2000), however, this binary of relativism versus objectivism was rejected under the assumption that truth was

tied to issues of chronology, economics, and aesthetics within the contexts of history, politics, and power. She proposed that postmodernism operated under the assumption that problems did not have generalizable solutions because they existed contextually. With this assumption laying the groundwork for the methods of this study, it was not the purpose of this study to generalize to other populations. Findings from this study informed changes in the curriculum, but the intent was to describe methods that were contextually grounded.

Pillow (2000) described a postmodernist methodological approach as making oneself available to intelligibility (knowing) through a rhizomatic discourse in order to determine where your position gave you the most freedom to work while using your most passionate approach. As she advised, I defined myself in terms of the theory rather than trying to clearly define the parameters of the theory, and shut out some modes of thinking so I could think in the way that I best thought. A postmodern approach welcomed the rabbit holes of thoughts and advice, those things often considered tangents to the big picture as we know it, as a means for considering truth as opposed to what is passed down as truth. As I attempted to design and carry out the study, allowing the rabbit holes to overcommit me allowed for more knowing. On knowledge, St. Pierre (2000) stated that postmodernism was characterized by a charge of criticizing the assumptions regarding what was and what was not knowledge. This was also accompanied by a substantial suspicion of knowledge as universally applicable (Bazzul, 2013: Lather, 1991, 2004a; Pillow, 2000; St. Pierre, 2000, 2002). It was this clearly outlined definition of the postmodern approach that aligned with questions of social justice.

The postmodern path to knowledge ran similarly to rhizomatic Nandina root systems. St. Pierre (2000) described postmodernist study as a rhizomatic citational trail. Some rhizomatic plants had rhizomatic roots above ground as did ferns. The rhizomes were exposed and available for all to criticize, which described a postmodern researcher's approach to social policy, its stated intentions, and the real outcomes and extensions sought to uncover through study (St. Pierre, 2000, 2002). Some rhizomes have been used to make desirable additions to our food such as ginger. These rhizomes were hunted, dug up, and used to give new meaning to foods that were already acceptable without the addition. Postmodernists have made such additions to improve the perspective of findings so that the uniqueness of the context was as complete as it could be (St. Pierre, 2002). Some rhizomes such as asparagus have been used as a food outright, requiring no collaboration as were some of the wonderful rabbit holes an undefined approach could afford. This rigorous confusion may have appeared to be chaos to those who love to live in the box or in the semi-box.

Crotty (2003) also defined the "post" in postmodernism as nothing of the chronological nature. Instead, postmodernism incorporated many of the questions asked by modernists. He affirmed that postmodernism differed from modernism in that relativity to context was the driving assumption. The two theories were not utter opposites. Postmodernists argued that the approach was both "subversive and redemptive" (p. 193) with respect to the social world we share; Ladson-Billings (1998) also proposed that critical theory, and thus postmodernism, belonged in education when the intent was to study the social construct of an educational setting in terms of how that

interacted with marginalized groups. That idea was applied to this study in the context of critical considerations of the setting in which females received physics instruction.

Critical Theory

The issues of gender threat and marginalization within the K–12 STEM experience have been reported in the literature extensively (Bilimoria, Joy, & Liang, 2008; Buse, Bilimoria, & Perelli, 2013; Servon & Visser, 2011). It has also been suggested that getting an education was a construct, with females having the perception that being serious about your school work was more characteristic of science majors (Eisenhart & Holland, 1990, p. 164). The extensive work of Eisenhart and Holland (1990) indicated that ideas about the reasons for getting an education were socially and culturally constructed, with highly capable females heavily weighing peer-relationships and romantic involvement against academics. These ideas served as a foundation for the development of this research. Though postmodern approaches have rejected structure or containment of methods during the process, there were rabbit holes that developed during the course of designing this dissertation work that created a need for an additional theoretical consideration. Critical theory has been used to study social justice issues of racism, classism, regionalism, sexism, and more (Carrell, Page, & West, 2010; Chambers, 2009; Gorski, 2008, 2010, 2011; Guiso, Monte, Sapienza, & Zingales, 2008; Gutiérrez, 2008; Ladson-Billings, 1998, 2006; Lather, 1991, 2004a, 2004b). As applied to this study, critical theory held that gender did not influence physics performance when

instructional decisions in teaching method and assessment were taken into account because the proposed deficits of females were socially constructed rather than real.

Subjectivities: Influential Coursework

During David Larimore's lectures for graduate students in quantitative research methods courses at TTU, an incredible number of instructor comments struck me as quotable. Often delivered with striking rhetoric for the purpose of making a point about research practices and interpretation of findings, I was unable to resist the urge to document the comments that took me to my experiences as a high school STEM teacher and to the research I was reading in the STEM research course taught by physics faculty. I included some of the quotes here because they were perspective-changing and directly influenced my positionality as well as methodology choices in this work. The comments from lectures were italicized and were documented in notes taken during class.

On the subject of scientific objectivity and the treatment of randomized controlled trials as a universality of positivistic designs, the use of rats to gain knowledge in psychology was later applied to easily attainable human subjects for research conducted at higher education institutions: *Everything that was true for rats turned out to be true for sophomores and then true for other people, too.* This comment was relevant to this dissertation work in that researchers worked with populations that were available to them, possibly influenced by the need to show merit of the research before being able to move on to the original population of interest. For the action research described here, students of introductory algebra-based physics at one university were studied. There was no

intention to say that the findings for these *rats* were true for all physics learners, nor for all STEM learners. These findings, however, informed efforts to study the learning of *sophomores* and possibly *other people*.

You were born on third base and think you hit a triple. Easily entertained by transgressive methods for knowing and representing information, this comment supported my critical position on gender differences research. The intent was to draw an analogy to the use of data to represent differences without methods for controlling for ability. If there were factors that predisposed a student or group to perform well, it was inappropriate to attribute performance to any other variables. My position on the study of differences was that differences in pretests should have been carefully considered to be a reflection of differences in mentoring and opportunities, rather than personal deficits, before publishing a work that could lead to unforgettable headlines. In this study, pretest performance and other covariates were accounted for so that inappropriate information was not used to make suggestions for future curriculum development and research.

On the subject of the perceived gender gap in STEM performance, I had a deep-seated need to know if many of the differences between groups were superficial and a product of the policies and practices that led up to the measurement of differences. The presence of powerblind discourses was most irritating, and this influenced my every decision. Because of my analytical and critical tendencies, as confirmed by Gallup Strengths Finder analysis conducted as a professional development activity at Middle Tennessee State University, I was attracted to disciplines where I thought I could most easily be an analytical thinker. When I was a girl, my perception was that science was most conducive to analysis. I also perceived that finding the discrepancies and faults in

matters—being critical—was inherent to science. I spent a good portion of my life ignoring my interest in social justice matters, entertaining myself with anthropology courses. It has been a quality-of-life-altering experience to find these often separated worlds operating together in my daily thinking—science and social justice. *Discovering the truth in all of the stuff that gets passed down as the truth was interesting to me.*

Research Assumptions

It was assumed that the sample was representative of the total population of students who enrolled in PHYS2010 at TTU. The demographics and background information obtained through the TTU Department of Enrollment Management and paper student records were free of error and described each student at the semester in which they were enrolled in PHYS2010. Assumptions of accuracy should trouble a postmodern, critical researcher. Within the positivist paradigm, responses to questions on conceptual instruments were assumed to accurately reflect the thinking (Lather, 2012), whether Newtonian or Aristotelian, of the student. St. Pierre (2000) would have said that it is impossible to know if the students' answers accurately reflected their thinking, as accuracy is a construct elusive to knowing.

Summary

The purpose of this work was to look at the ways pedagogy and assessment choices may have led to a gender gap that was not real. Education research designed

using deficit ideologies may have further marginalized target groups and produced headlines that put undue focus on 'fixing' those people who simply have experienced deficits in access and opportunities. Here, schooling was considered to be subject to the culture of the individual classroom while the focus was on the amount of learning that took place with respect to the pedagogical choice—without further consideration of background and demographics once they were taken into account. The focus was to evaluate learning using a statistical approach that could be described as strong objectivity (Harding, 1993a). To accomplish these goals, a probability (i.e. propensity score) of choosing LEAP instead of a traditional section of PHYS2010 was generated for each student to attempt to equate the self-selected groups. This variable was a conglomerate of many background and demographic covariates. Differences in physics performance on the FCI and GFCI were then evaluated by pedagogy and gender while blocking on the probability of choosing LEAP and pretest performance.

The remainder of the study is organized into four chapters, a list of references, bibliography, and appendix in the following manner. Chapter 2 begins with a warning of the implications of deficit models and then presents literature addressing gender issues as related to physics education and pedagogy discourse. An introduction to FCI and GFCI studies relevant to this work follows a description of the development of those instruments. Validity and reliability of each instrument is presented in Chapter 3. Chapter 3 also describes the utility of propensity scoring as a pre-analysis technique for dealing with self-selected groupings in a quantitative evaluation of group differences and concludes with a description of the variables used for each research question in analysis of variance (ANOVA). Results of propensity scoring and its implications for FCI and

GFCI analysis are presented in Chapter 4, as well as a disaggregation of performance based on constructs of the FCI. In Chapter 5 a discussion of findings pertinent to the comparison of LEAP and traditional pedagogies is presented, gender gap discourse is revisited in the context of the findings, and a critical perspective on what should be done in light of the findings is offered.

CHAPTER 2

REVIEW OF THE LITERATURE

The literature presented here was used deductively to establish the need for a critical framework and informed the research questions. A review of deficit theory was presented within the contexts of achievement and gender. A review of this aggressively critical perspective on educational achievement discourse was necessary in order to provide transparency to this research agenda. Following the critical considerations of the dominant gender and achievement discourse, relevant literature on STEM education was presented. Literature on signature pedagogies in STEM provided background on one of the two major independent variables of the study. The FCI and GFCI assessments were described, including development of the instruments.

Deficit Ideology

This study was conducted through the lens of postmodernism and critical theory. Gorski (2011) distinguished between focusing on difference as opposed to equating difference—from ourselves—to deficit. Within a deficit ideology, the larger social, political, and economic contexts (within which schooling is situated) has been considered unrelated to the academic performance differences between individuals and communities (Bensimon, 2005; Bomer, Dworin, May, & Semingson, 2008; Chambers, 2009; Dudley-Marling, 2007; Gorski, 2008, 2010, 2011; Gutiérrez, 2008; Ladson-Billings, 2006).

Rather, characteristics drawn from the popular stereotype for a group were used to explain and support a deficit theory (Gorski, 2011). Deficit thinking has become the dominant discourse and has been maintained by marginalizing any discussions to the contrary (Gorski, 2008, 2011). Gorski (2011) proposed that this marginalization occurred through the use of stereotypical images to propagate deficit thinking and normalize people to, or assimilate people to, the overarching assumption that education was equitable by design and provided opportunities for all. Bensimon (2005) also pointed out that deficit thinking was not necessarily expressed in negative ways, but genuinely concerned people may have still responded to the underperformance of a group of students by attributing low performance to stereotypical characteristics. Germane to gender studies was Gorski's (2011) observation that all of the misplaced urgency in addressing "achievement gaps" that were measured by unquestioned but "standardized" tests was a symptom of well-propagated deficit ideology. Gorksi's take on class and race differences, which he viewed outside of a deficit model, easily translated to issues of gender equity.

The Deficit Model of Achievement

The differences in performance among and between groups have been described as the *achievement gap* when defined within the deficit paradigm. As a rebuttal to deficit ideology, others have criticized the deficit-focused terminology by offering critical alternatives to describe the phenomenon such as "education debt" (Ladson-Billings,

2006), "the receivement gap" (Chambers, 2009), and the "gap-gazing fetish" (Gutiérrez, 2008). Scarce was the educational headline without a reference to the "achievement gap" or some similar alarmist technique for promoting intervention or reform. Even in 1877, teachers negotiated merit-based reprisals by encouraging students to leave school before testing (Tyack, 1974). Lather (2005, p. 2) attributed this "rage for accountability" to the consensus that objectivist approaches to science education research were best. Ladson-Billings (2006) illuminated the need to critically consider such tactics by drawing a metaphor between the national deficit and the achievement gap. Through a postmodern lens, she allowed a rhizomatic relationship between the nation's economic debts to take a line of flight to education debts. She troubled the cultural deficit ideas of achievement gaps by equating the cumulative effects of national budget deficits on national debt to the cumulative effect of each year's misappropriated resources on the differences between groups. The achievement gap, like the money gap, was created by yearly deficits and can only be narrowed by reducing the education debt of the past. She pointed out there were no clear reasons for why the achievement gap fluctuated, but there were clear reasons for the education debt. Like Gorski (2011), Ladson-Billings cited the attributing of stereotypes of a group to gaps in performance as a litmus test for deficit ideology.

The concept of the achievement gap was metaphorical to accumulated debt in Ladson-Billings' (2006) perspective on the differences in education outcomes. Chambers (2009) resonated Ladson-Billings' (2006) theme of misappropriated resources and subsequent accumulation of debt owed to students; Chambers (2009) suggested looking at differences through a perspective of gaps in *receivement* rather than gaps in achievement. She suggested that performance of minority students has been distorted by

a student-output focus and may have been more relevant and equitable if addressed from a structure-input focus for transformation towards a multicultural education. In her qualitative study of seven black students from a diverse high school, Chambers (2009) described the cumulative effect of ability placement interventions in elementary school on assumptions-based tracking at the high school level. Though she looked at the stereotypes assigned to young children and the observable relation to race, her study supported the theme of my research in that deficit thinking led to early interventions meant to fix the child as opposed to fixing barriers to early and continual access to opportunity.

Gutiérrez (2008) warned that focusing on gaps in achievement created a narrow view of equity issues and solutions. Deficit ideology has been perpetuated and supported by gap-focused discourse (Gutiérrez, 2008). Termed *gap-gazing*, she suggested that researching for the purpose of determining gaps, as well as efforts to ascertain how factors add to gaps, will not contribute to equity for marginalized groups. Bensimon (2005) pointed out that disaggregating student outcomes to evaluate the progress of groups of people, rather than the entire population being studied, was not standard practice in higher education. For those institutions of higher education that have investigated the progress of underrepresented groups, the purpose was likely for studying the diversity of the institution as a characteristic (Bensimon, 2005). Outcomes of institutional studies described in this manner were more likely to be explained with a deficit ideology thus perpetuating stereotypes and creating inequity (Bensimon, 2005). Through her organizational learning cognitive frame, a school that handled underperformance using a deficit approach was recognizable by the presence of

remediation measures and programs that focused on fixing the student. For the LEAP program at TTU, the deficit was proposed to be located in the ability of curriculum and pedagogy to promote understanding—not an individual deficit. Strong ties to this perspective were demonstrated by continual revision of curricular components and pedagogical awareness.

Synthesis of Deficit/Gap Literature

This study was situated in the critical discourse of Chambers (2009), Gutiérrez (2008), Ladson-Billings (2006), and Lather (2012) in that an achievement gap focus was a deficit model that further promoted stereotypes, placed the burden of equity on the marginalized, and allowed for accumulation of inequity. To support the idea that gaps were a symptom of an accumulation of individual inequities of the past as well as inherited generational inequity, my research focused on perceived gaps as a function of the way performance was assessed and the way data was interpreted. The entire research process was conducted with a perspective similar to that of Paul Gorski. In short, this study sought to determine if the gap was real or simply an outcome of inequitable pedagogical approaches, assessment instruments, and deficit thinking.

Females, Girls, Ladies, Women & Schooling

Though many groups of people fought and won great advances for minorities early in the history of the United States, these advances were not without a price. Many groups traded religious freedom and the right to practice non-Protestant cultural traditions for an education disguised as public assistance (Spring, 2004). Though this education was not equal in any way to the education being afforded to privileged white males, minorities began to be educated as early as the colonization of America (Tozer, Violas, & Senese, 2002). An educational system was built for the purpose of assimilating Native Americans posing resistance to the westward spread of White Protestants as well as those groups threatening capitalism and the accumulation of property by the Protestant political majority (Spring, 2004). The cultural and linguistic genocide of Native Americans discussed by Spring was a forecast of similar efforts to educate other minority groups for calculated, oppressive agendas. Spring (2004) described the differences in Africans, Asians, and Native Americans as counterproductive to the politically driven education movement designed to produce a productive, silent middle and lower class and a productive, powerful upper class. The desired result was a homogenized nation that would produce conformed, socially obedient hierarchies of people (McClaren, 2003).

The education of females did not respectably begin until many other minorities had fought and won the right to be educated. Thomas Jefferson felt education and freedom was deserved by all men simply because they were uncultivated potential capable of producing self-sustaining families if educated to an appropriate extent (Tozer et al., 2002). The authors described Jefferson's view of female education as necessary

only to the extent that allowed them to be mothers, homemakers, and educators of daughters. The curriculum provided to girls indicated that the ability of females was thought to be spent at the elementary school level and accurately reflected the belief that females were less educable than black males, Native American males, and especially white males.

Aside from the theory that females were not educable past elementary school, one had to consider that females served a role in the family that would have caused problems if replaced by a pursuit of intellectual understanding. Who would have cleaned, cooked, tended to babies, and attended to the "household economics" that Jefferson spoke of when justifying his lack of attention to the education of females (Tozer et al., 2002, p. 40)? The self-sustainable family unit described would not have been possible if women pursued something other than marriage and the home. Girls were taught at home and only if wealthy enough to have a literate mother or a private tutor. During the late 18th century, girls attended schools in the summer while boys assisted with the family farm. Girls were not educated with boys since the goals for female education were not the same as for males. The education of females was merely necessary at this time because moral examples and the teaching of children until they entered school was part of the goal for a unified, conformed nation. Prominent female advocates for the education of females felt an education for girls was to be a preparation for a future as a mother and wife which led to the "Cult of Domesticity" (Tozer et al., 2002, p. 128). The major goals of the curriculum were to develop women who were better domesticators and companions to their husbands. Though they were advocating for women, the view was not one of equality.

The English were successful in creating stable family units through domestic training for women and assimilation through education (Robenstine, 1992). Robenstine's (1992) historical account of French colonial policy described the French colony as dominated by single males, adventurous but unstable types, and minorities. The French were politically motivated to evolve the community into a French replica of the English success (Robenstine, 1992). Without coherent families the colony would not grow or become stable, thus creation of family units for the purpose of developing stable and rooted French colonies was a priority (Robenstine, 1992). In order to draw marriageable women into the colony, the Company of the Indes contracted French nuns to bring instruction to all girls in the colony, regardless of class or race (Robenstine, 1992). In Robenstine's account women were considered the stable sex capable of maintaining the family unit, thus their education was for the good of the group. The use of religious school as the educating institution was a religious matter on the surface; the reality was the underlying use of education as a political tool (Robenstein, 1992). There was no sincere concern for minority equality in French or English colonial education agendas.

Feminization of education was a result of the woman's role in the family at the time. Horace Mann proposed that nurturers were needed as teachers and that females were better equipped to fill the role of compassion (Tozer et al., 2002). The suggestion that women would be good teachers of children was premised with the idea that women could work for less, relieving males to work jobs that served the nation and properly utilized their capabilities (Tozer et al., 2002). Teachers were hired barely literate because they would accept an inadequate salary; with few other markets competing for the female labor force, any pay was better than no pay (Tozer et al., 2002). In some urban areas

teachers were brought together for curricular meetings in which a male administrator went through textbooks page by page while telling them what questions to ask for each lesson, instructing that there was to be no deviation (Tyack, 1974). It is arguably true that teachers being mostly female, and peddling Protestant values, limited the diversity of experiences offered to all students, especially girls. This result served well the objectives of education during that time.

The education of females today has not completely evolved from the type of education provided to females at the time of the Revolution. Many current practices have reproduced social hierarchies, particularly gender-biased policies and practices (Asher, 2002; Britzman, 1997; Lather, 2012; Schwalbe et al., 2000). Eisenhart and Holland (1990) followed highly capable women throughout their college experience to determine what was at play in decision making and education constructs. Many of the women felt the purpose of education was to attain a degree, rather than learning for a life's work (Eisenhart & Holland, 1990). The outcomes of that qualitative work, titled *Educated in Romance*, was a sobering account of how engendered the construct of education was, even recently.

The subordinance within schooling has been discussed as particularly weakening for girls being taught by women. Even more daunting was the potential for sexism introduced through homogenizing efforts within a nation. Female students were taught feminine social roles through the subservient nature in which educators accepted testing and standardization initiatives without question or at least without action. During the industrialization of America, workers were deskilled which allowed lower wages and easily dispensable employees (Tozer et al., 2002). The introduction of standards and

curriculum frameworks arguably deskilled educators whose talents exceeded or fell outside of the specific subject matter placed on standardized tests. Such deskilling of educators could have created a more homogeneous set of teachers, and the message to students was that women were not capable of developing their own agenda in the classroom. Girls have been bombarded daily with images of the teacher being compliant and subservient. The Cult of Domesticity has been perpetuated.

If the socially constructed deskilling and hegemonic practices of educated female teachers was not enough to concretize gender stereotypes, stereotyping in teaching materials was another profound source of hidden agenda sure to produce a gap. The depiction of women performing menial jobs that typically offered poor benefits was a unifying theme of textbook stereotype research and was supported by a large international body of studies synthesized by the Equal Opportunity Commission of Hong Kong (2001). These depictions implied low expectations of females, and emphasized the role of marriage as a way to escape the substandard options displayed in text content (Shore, 2000). Studies analyzing the role portrayal of texts found that females were portrayed as observers and agreeable onlookers, rather than executors of skilled tasks (Commeyras & Alvermann, 1996). Stereotyping such as this contributed to the concept of girls as homemakers or public assistants such as clerks and caregivers. Roles accepted as teachers, as well as the practices adopted, have set examples for girls. Our textbooks have been nothing more than technical counterparts to romance novels if such stereotyping existed (Shore, 2000).

As spokesperson for the America Foundation for the Blind, Helen Keller was advised that her activism was a source of embarrassment, and there was no discussion of

her activism against capitalism and social injustice found in school materials, simply an account of accomplishments thought to be courageous for the blind, such as riding a bicycle (Hubbard, 2003). Helen Keller was depicted as famous because of her willingness to help others, not because she rose up against the social injustice she found around her (Hubbard, 2003). Reflecting upon the helping nature depicted by the females found in textbooks and the reality that female interests were situated in helping-careers, there should have been little doubt that gender roles were defined by such teaching materials.

The Cult of Domestication has not gone away, it has simply resided in the bias we have allowed to go unquestioned in our classrooms. Female education was never intended as career preparation, but rather preparation for being loyal homemakers. We cannot claim a change has occurred until we have questioned our purpose and influence as women educating girls. It is up to teachers to create an atmosphere that celebrates differences and questions the limits placed on our girls.

History of STEM Education

Following concerns about Americans receiving fewer science and engineering degrees than other nations and the less than competitive performances on international assessments (National Research Council, 2006), the National Science Foundation responded with a call for policies that supported an increase of minority representation in STEM majors and careers. A longitudinal study which oversampled minorities to identify constructs of minority persistence found that parent and high school mentor influence as a motivator to major in engineering was correlated to persistence, and confidence in math

and science skills was correlated to persistence (Eris et al., 2010). This has placed a large portion of the burden to produce enough engineers to sustain the country upon the backs of PK–12 educators and parents.

The current status of STEM education was characterized by a silo approach to teaching science, technology, and mathematics (Sanders, 2006, 2009). In this STEM experience, design of technologies was rarely taught in science class, and science standards were rarely taught in technology courses (Sanders, 2009). The changes at the Chalons school reflected what has happened in schools today as we attempted integration of technology education and engineering processes. As apprentice programs became rare, the French government implemented technology education in the schools of the 1790s (Pannabecker, 2002). Pannabecker (2002) described that schools experienced difficulties when integrating subjects, as has been experienced recently in our schools each time a new initiative created a need for redesign of curricula. Trends in integrative versus silo approaches to instruction were evidenced by mathematics being used as a term to describe math, chemistry, and physics courses, while teachers of the sciences were regarded as math teachers (Pannabecker, 2002). By the 1820s, it seemed that Napolean's schools had removed integration for a more silo-like approach (Pannabecker, 2002). Schools of today have wrestled with something similar in response to having students of varying abilities and interests. The school of Chalons added new courses for the low-ability or younger students, while there were courses that required higher levels of content understanding and in which quality and precision were important for the nation (Pannabecker, 2002). The American Association for the Advancement of Science (AAAS) advised that the facets of technology, as well as responsible analysis of the

impact of the development and use of a technology, were fundamentally necessary for becoming scientifically literate (Rutherford & Ahlgred, 1989). The AAAS cornered the idea for making these areas of study interdependent at least two decades before the national and state standards reflected this integrative perspective (International Technology Education Association, 2000; National Research Council, 2012; Rutherford & Ahlgred, 1989). Without experiences in integrative STEM learning, students had little experience as a STEM worker in the real world (Carlson & Kwon, 2006; Sanders, 2009). New approaches to integrating STEM education were of interest to educators, business, and government for the purpose of improving the diversity and competitiveness of the STEM workforce (Sanders, 2006). A disconnect between a silo STEM education experience and the intensely integrated STEM workforce has not improved the lack of minority representation in STEM fields.

Gender & STEM Education

A conversation about the differences in performance of males and females in STEM should be started and cultivated by women even if it will go unheard or ostracize the speaker (Asher, 2002). At the head of the argument for a close look at the structures that set the stage for inequities was an alarming difference between the number of college graduates who work and are women (half) and the number of STEM degree recipients who work in STEM fields and are women (20%), with women who have STEM degrees being more likely to work as teachers and healthcare providers (Beede et al., 2011). For

women who have STEM degrees, only 26% worked in STEM fields as compared to 40% of men who have STEM degrees (Beede et al., 2011).

Underrepresentation of females in STEM majors and careers has been attributed to biology (Baron-Cohen, 2007; Kimura, 2007), education structures (Hines, 2007; Hoffman, 2002; Zohar & Bronshtein, 2005), and culture (Baram-Tsabari & Yarden, 2011; Hewlett, Luce, & Servon, 2008; Hines, 2007; Kelly, 1978; Spelke & Grace, 2007). The biology debate was supported by the theory that hormones which influenced spatial reasoning, particularly androgens, also mediated interest in science (Kimura, 2007). Hines (2007) argued that spatial reasoning was not related to levels of male hormones. The notion that females were motivated towards fields that involved helping people was attributed to contradictory variables of biology (Baron-Cohen, 2007) and cultural expectations (Baram-Tsabari & Yarden, 2011; Guiso et al., 2008; Spelke & Grace, 2007). Having a female teacher was shown to increase math and science performance of females, especially for females who performed in the upper 5% of the nation's student population; having a female teacher was related to increased enrollment in higher level math and science courses as well as STEM degree attainment of female students (Carrell et al., 2010). Performance of male students was not affected by the gender of the teacher (Carrell et al., 2010). Women who persisted in engineering careers, rather than having left once there, tended to have fewer children and were willing to navigate discriminating encounters rather than submit through silence (Buse et al., 2013). These findings supported the notion that being educated as a female in STEM was a cultural construct of its own, that might be useful in looking at gender issues in STEM.

Gender & Mathematics Education

Just as mathematics was used as an umbrella term to describe math, chemistry, and physics courses, while teachers of the sciences were regarded as math teachers (Pannabecker, 2002), mathematics was presented in this study in the context of gender as if physics teachers were teachers of mathematics. Germaine to the topic of STEM education and gender was the accumulating lack of access to mathematics education (Confrey & Lachance, 2000). Even if the prevalence of failure in mathematics throughout the population was set aside for a gender discussion, structures of schooling and teaching methodologies were suspect in the search for reasons for poor performance (Chambers, 2009; Confrey & Lachance, 2000). Minorities and females were less likely to be enrolled in high level mathematics courses, an outcome of school ability-tracking structures which catalyzed accumulation of inequity (Confrey & Lachance, 2000; Gutiérrez, 2008). Confrey and Lachance (2000) pointed out that "we are supporting an elaborate mathematics education system that succeeds for only a tiny percentage of the population" (p. 233).

Meta-analysis of gender and math performance showed little difference between males and females as early as the 1970s (Hyde, Fennema, & Lamon, 1990). At that time, small but significant differences between male and female problem-solving performance existed at the high school level (Hyde et al., 1990). A lack of high level math coursework was equated to the differences between males and females (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). No gender differences at any level of mathematics education were found in a 2008 meta-analysis of state standardized test performance though authors

stipulated that none of the state assessments included problem-solving (Hyde et al., 2008). Hyde et al. (2008) further purported that the anxiety-driven choice to center instruction around high-stakes state standardized tests meant that problem-solving skills may have been getting benched in high school. If this was true, the college STEM experience could very well be the initiation into the problem-solving skillset needed in STEM majors and careers, thus disenfranchising students who were otherwise interested in STEM.

In a cross-national meta-analysis of gender differences in mathematics performance on standardized tests, no statistically significant differences between males and females were found (Else-Quest et al., 2010). Males of all nations were found to have better attitudes towards mathematics (Else-Quest et al., 2010). Great variation in gender performance differences was found between nations and was related to the social status of women, school enrollment equity, and women holding research careers (Else-Quest et al., 2010). Guiso et al. (2008) also found that performance in mathematics was tied to the gender equity of a nation, with no performance differences found in nations characterized by gender equality.

Gender & Science Education

Differences in science performance of males and females have been attributed to culture (Baram-Tsabari & Yarden, 2011; Hewlett et al., 2008; Kelly, 1978), attitude (Osborne, Simon, & Collins, 2003), and educational structures (Hoffman, 2002; Zohar & Bronshtein, 2005). Girls identified themselves as science persons less than boys, though

performance in science was similar (Barton, Tan, & Rivet, 2008). There was a hidden curriculum in which girls were led to believe that a scientific identity was not conducive to the gender identity (Barton et al., 2008). Interest in science became increasingly different for males and females at the middle school transition and again at the transition between tenth and twelfth grade, though interests did not differ in early elementary years (Baram-Tsabari & Yarden, 2011). Barton et al. (2008) suggested that authentic science practices included both learning the content and learning how to participate in the science field community. Hybridity provided the opportunity for girls to "gain epistemic authority in the classroom" (p. 74) and thus overcome barriers inherent to the subject area (Barton et al., 2008).

By high school—and markedly so in college—females preferred biological studies while males preferred physics and technology (Baram-Tsabari & Yarden, 2011). Though females earned more degrees than males, males earned more science degrees than females (Beede et al., 2011). Women believed they would be marginalized in the culture of the science workplace due to male-oriented norms for work habits (Servon & Visser, 2011). To increase the number of women in science who stay, Bilimoria et al. (2008) suggested that a deliberate restructure of the culture of academia and the workplace was necessary to "break down the barriers constraining women's participation and effectiveness" (p. 423).

Gender & Physics Education

Gender differences in physics interest have been found to increase with age (Baram-Tsabari & Yarden, 2011). The masculine stereotype more prevalent in physics than in other sciences (Kelly, 1978) was thought to contribute to the increase of gender differences in science interests as a student grew older (Baram-Tsabari & Yarden, 2011). Waning interest in physics has been attributed to the underrepresentation of females in physics courses and related careers (Baram-Tsabari & Yarden, 2009; Krapp, 2000). Traditional physics pedagogy and gender-biased assessments and curriculum have been linked to interests in physics (Hoffman, 2002; Zohar & Bronshtein, 2005). Changing contexts to a female-oriented scenario has been proposed as a way to increase interest in physics and reduce negative experiences that stem from continually encountering scenarios that are unfamiliar (Baram-Tsabari & Yarden, 2008.)

While females enrolled in biology and chemistry more than males, males continued to enroll in physics more than females (Zohar & Sela, 2003). Physics enrollment of females was related to the types of jobs held by women in the community (Reigle-Crumb & Moore, 2014). For communities with more females in STEM jobs, physics enrollment of females was greater (Reigle-Crumb & Moore, 2014). By treating gender as a socially constructed variable, these researchers supported the notion that being female was a culture of its own. When comparing countries by which area of science showed the greatest gender performance difference, physics had the largest difference in the most countries (Zohar & Sela, 2003).

When looking at background characteristics to explain the differences in performance between males and females, prior physics courses and performance in math have not explained those differences (Kost et al., 2009; McCullough, 2002). In a study using matched samples to look at differences in performance between males and females while attempting to control variables of high school physics enrollment, high school GPA, last high school mathematics course taken, locus of control over grades, year in college, FCI pretest score, a free response conceptual pretest score, and a problem-solving test, no statistically significant difference between males and females on the posttest was found (Blue & Heller, 2003). Blue & Heller (2003) found that when differences brought to class were controlled for, males and females learned the same amount of physics. They concluded that the differences in performance of males and females were due to the differences in the culture rather than being physically male or female (Blue & Heller, 2003). Sabella and Van Duzor (2013) suggested that tapping the cultural capital brought to the classroom by students, rather than choosing to focus on the present culture of the discipline, was a means for adjusting the culture of science to better accommodate for all students.

Of the student characteristics looked at in a study of physics students at a Canadian university, gender was most related to performance on the FCI pretest (Noack, Antimirova, & Milner-Bolotin, 2009). The authors looked at coursework, grades, and demographics to determine predictors of performance. Completing an upper level physics course in high school explained the greatest portion of the variance in pretest performance between males and females compared to other educational characteristics that were studied. However, the educational characteristics and demographics included in

the study were not related to performance gains on the FCI. The authors warned introductory physics instructors about "substantial limitations in terms of the goals that first-year physics instructors can expect" (p. 1274). The results of that study provided support for looking at the context of the setting in which physics learning took place. A look at where students lived prior to joining the community in which learning takes place provided insight into the context for learning of my study.

Pedagogy & Pedagogical Knowledge

The National Research Council (2002) publication, *How People Learn*, was a response to the need for a marriage between cognitive science, teaching practices, and pre-service teacher education. A unifying theme of the text was that misconceptions built upon ill-conceived personal theories of how things happen did not provide a stable foundation for building new knowledge. Farnham-Diggory (1994) conducted a review of three instructional paradigms that she identified as three "core" methods of instructing and thus theories of transformation from a novice to an expert. She posited that any educating that occurs can be classified into one of the three paradigms. The borders between these three paradigms were defined by 1) the thinking that one used to distinguish a novice from an expert and 2) the "mechanism" by which a novice was developed into an expert (p. 464).

In the behavior model, education was an empirically studied phenomenon since all outcomes were measurable (Farnham-Diggory, 1994). Thorndike inspired this transmission model of learning. Working under the premise that what is taught, how

often it is reviewed, and so on can be measured, it seemed reasonable to visualize drill and practice as a method for transmission of information to the novice. In this paradigm, a learner's progress could be visualized as a number line in which each graduation represented a percent of what constituted knowing for the outcome. As the learner proceeded along the straight path from novice towards expert, it was assumed that the process occurred by gaining "increments" from the expert—much like tossing information in until it remained there (p. 465).

The apprenticeship model was, unlike the others described, not "culture free" (Farnham-Diggory, 1994, p. 466). The assumption that learning as well as progression from novice to expert only happened when acculturation took place was the hallmark of the apprenticeship paradigm. This would only be possible if learning took place by observing, accompanying, and then demonstrating independent competence at the task. To know, one must have belonged. The novice first was an outsider void of knowledge of content and norms of "intellectual allegiance" (p. 466). As the beginner was assimilated by practicing alongside experts, independent application of tacit theories demonstrated that experience within the culture informed intermediate thought. The intermediate practitioner became an expert when capable of fully independent knowledge-making and application.

Piaget laid the groundwork for the developmental model of learning and transformation from novice to expert (Farnham-Diggory, 1994). The mechanism by which a learner gained personal theories that he or she used to explain phenomena or account for observations was considered to be the target of change used by educators living in this paradigm. An educator provided "perturbing" events which served to

discount unsupported personal theories, elicited new explanations, and revised the means by which the learner explained phenomena (p. 465). This global shift in thinking was akin to the constructivist process, if they were not synonymous. The role of the teacher was to elicit misconceptions and the underlying ill-perceived theories as a means for customizing activities that served as perturbing events. After experiencing these activities, the gain was not intended to be quantitative. A student began to adjust his or her theory on a phenomenon or process when provided a "perturbation" of "the student's personal theory" (p. 465). The perturbing event provided by the expert caused interaction between the novice and personal beliefs of the novice.

The learning environment experienced by a learner determined what was known and how it got known (Goldberg et al., 2010). The physical body, the culture in the classroom, the beliefs, and prior knowledge of the student must be considered as contexts of learning (National Research Council, 2002). According to Hake (2007), education research should have focused on undergraduate education, since the means by which undergraduate students received their educations is the very means by which they have educated others, whether elementary students, another generation of undergraduates, or employees (Hake, 2007). Hake (2007) made clear that professors ought not to have grumbled over the unprepared student which was taught by the PK–12 workforce trained by the same generation of professors.

Just as law students would not speak to each other directly when presenting a case to the instructor, science students working on an experiment would not communicate scientific ideas through a third party judge. This example illustrated Shulman's (2005) idea that *signature pedagogies* were those that allowed students to conduct themselves as

members of the profession while in the classroom. Since the student discourse and "the physical layout of classrooms so typically tracks the premises of a field's signature pedagogies, the very architecture of teaching encourages pedagogical inertia" (Shulman, 2005, p. 57). Work as a scientist has often looked nothing like students of science doing classwork. Scientists have rarely conducted themselves as so by remaining in a chair facing a lector, copying the thoughts of an authority onto paper, or following a set of instructions to conduct an experiment. The dominant, traditional instructional paradigm (transmission) in science education has not modeled the practices (developmental) of science professions.

Traditional Pedagogies

Traditional physics courses were defined in this study as teacher-centered courses, in which students had a passive role, characterized by lectures, recipe laboratory activities, and algorithmic exams (Hake, 1998). Regardless of the intellect, abilities, or student ratings of the instructor, traditional pedagogies have done little to increase student conceptions of Newtonian phenomena (McDermott & Redish, 1999). According to Hake (1998), traditional passive-student introductory physics courses were characterized by the absence of interactive-engagement and low conceptual understanding even after instruction by favored instructors. For the traditional classrooms defined in this study, the lecturer was considered to be the deliverer of knowledge.

In a study of physics students' learning in a highly competitive, high stakes testing environment, Zohar and Sela (2003) interviewed females to elicit perceptions of

the learning environment. Teaching pedagogy was greatly criticized by the female participants, with traditional transmission, lecture, and focus on algorithmic processes questioned in favor of deeper understanding. Zohar and Sela pointed out that the success of girls in physics was a remarkable accomplishment considering that teaching methods often were not equitable. The assessments used to measure these students were criticized for lacking the questions that assessed deep understanding rather than algorithmic processes, which led to test refinement and was noted as possibly related to later increase in female performance (Zohar & Sela, 2003). This was significant to the need for refinement of the FCI in light of contradictory evidence for the correlation of background to performance on the FCI.

Cronin Jones (2003) argued that some material was best suited to a lecture venue. However, the lecture she described was simply lecture-hall interactive engagement conducted by a lector whose plans were informed by the constructivist paradigm. In the seminal pedagogical piece *A Time for Telling*, Schwartz and Bransford (1998) found that lecture transmission of information was an important part of a constructivist pedagogy when well-timed with student knowledge. Said another way, telling worked when relevance was had by the listener. Schwartz and Bransford (1998) also went so far as to point out that engaging in inquiry alone did not lead to learning. The opportunity for telling described in their study was situated in the constructivist paradigm and should not be mistaken for the continual talking of a lecturer using a traditional pedagogy.

Reformed Teaching Pedagogy

For the purposes of this study, the constructivist paradigm was situated in a socio-cultural view as well as the socio-linguistic constructivism stance of Vygotsky (Pibern & Sawada, 2000). Teaching behaviors that demonstrated constructivist pedagogies within a student-centered interactive engagement classroom environment were defined here as reformed teaching (Adamson et al., 2003; Goldberg et al., 2010; MacIsaac & Falconer, 2002; Morrell, Flick, & Wainwright, 2004; Piburn & Sawada, 2000; Sawada et al., 2002; Wainwright, Flick, & Morrell, 2003; Wainwright, Flick, Morrell, & Schepige, 2003). Physics courses with interactive engagement pedagogies were defined here as student-centered courses in which heads-on and sometimes hands-on work was done throughout the lesson for the purpose of yielding immediate feedback from the instructor (Hake, 1998). This immediate feedback from the instructor came in the form of face-to-face query as well as the whole-class telling that followed discrepant cases or other signs of readiness (Schwartz & Bransford, 1998). Tools were utilized in ways that allowed students to practice manipulation of science and math tools in ways that a practitioner would (Goldberg et al., 2010).

The PET curriculum was informed by these tenants of reformed teaching and cognitive science (Goldberg et al., 2010). The PET curriculum incorporated the norms of peer talk, evidence-based defense of ideas, and retention of personal ideas that were supported by evidence generated through experimentation. Interactive engagement strategies elicited academic peer talk, disturbed thinking built upon misconceptions, and provided continual feedback that was immediate (Goldberg et al., 2010; Hake, 1998).

Through this peer talk the responsibility of sharing your disagreement with peers was taken seriously, because that was how scientists worked (Shulman, 2005). There was expected to be member checking with respect to whether or not procedure was going as directed. Students developed the norm of thinking aloud and responding in the form of a question as a means of laying uncertainties on the table in an environment in which being wrong was simply a momentary position to be reflected upon once rectified (Goldberg et al., 2010).

Hake (1998) also found that interactive engagement courses produced higher average normalized gains, sometimes referred to as *Hake gain*, on the FCI. In that study, Hake solicited FCI data from high school and college teachers who used the FCI in their own action research, with teachers of sixty-two courses and over six thousand students included in the data analysis. Though an impressively large sample, Hake suggested that the course sample more heavily represented courses with greater student gains. Courses with minimal gains were less likely to be voluntarily submitted by teachers. More relevant to the current study was the fact that none of the sixty-two courses in Hake's (1998) analysis showed high gains ($\langle g \rangle \geq 0.7$). Aristotelian thinkers were identified by lower scores on the FCI, as a low score indicated that thinking was solidly grounded in that which went against nature (Hestenes, Wells, & Swackhamer, 1992).

Force Concept Inventory

Force Concept Inventory Development

The FCI was a diagnostic Newtonian–Aristotelian spectrum test developed by Hestenes, Wells, and Swackhamer (1992). The purpose of creating the FCI was to develop a diagnostic instrument that was an improvement on the Mechanics Diagnostic (MD) instrument, also created by FCI developers (Hestenes et al., 1992). It was not written for someone familiar with physics jargon. Rather than an achievement measure, this diagnostic test provided a description of the type of thought being used to explain a phenomenon in which force concepts should be applied. The FCI measured beliefs rather than intelligence or achievement. Commonsense beliefs about forces were incompatible with reality. The test developers took the position that commonsense beliefs should be treated as respected ideas that are "grounded in everyday experience" (p. 142). Newtonian mechanics revolved around force, so the FCI items were designed to force a choice between commonsense and Newtonian thought.

The concept of force was broken down into the six Newtonian dimensions. Half of the 30 FCI questions were taken from the MD. The conceptual items of the FCI revealed poor understanding though the test appeared trivial until results inevitably showed little change with a year of physics instruction. A premise of traditional test analysis was that a low average on a question could mean high difficulty or low ability (Wang & Bao, 2010). In the case of the FCI, it was considered to produce low scores on a question when misconceptions were present (Hestenes et al., 1992). The FCI content was

narrowly focused in comparison to physics content overall, but the FCI was not unidimensional (Wang & Bao, 2010). The item response model was used to describe features of the test that would not change if students changed, and thus provided the following properties of the FCI: high difficulty level, highly discriminant, and insignificant chance of correct guessing (Wang & Bao, 2010).

Field Testing of the FCI

The FCI was field tested with over 1,500 high school physics students from twenty Arizona high schools and over 500 college physics students. Except for two test developers, teachers of field tested students had no knowledge of the test design and were participants in the same physics pedagogy professional development summer training following their first year of participating in the FCI field testing.

Pretest and posttest scores from both the FCI and Mechanics Diagnostic were collected for all student participants (Hestenes et al., 1992). Students were given the Mechanics Baseline (MB) as a posttest to establish validity. Students from Harvard physics courses were administered the FCI and MB on computers for the purpose of determining how much time was necessary to ensure that student scores were not affected by time constraints. An algebraic test was used to describe student groups. The MB was not correlated to the algebraic test, except for one class of students. For this reason, math ability was concluded to be unrelated to FCI performance.

FCI developers interviewed twenty students from the high school and undergraduate participant pool about their responses to FCI questions (Hestenes et al.,

1992). For FCI questions that were answered incorrectly by the student, explanations were classified as Newtonian reasoning or non-Newtonian reasoning. The purpose of the interviews was to determine if the distractors for questions were working as distractors. This also gave insight into the potential for each question to differentiate Newtonian from non-Newtonian reasoning. Developers were also informed about common themes for misunderstanding across interviewees.

Interviews were also conducted with sixteen new graduate mechanics students for the purpose of eliciting reasons for their responses (Hestenes et al., 1992). Students who scored high on the FCI explained their choices using Newtonian-based reasoning while graduate students with the poorest performance on the FCI were also doing poorly in their graduate studies and exhibited non-Newtonian reasoning when explaining their FCI responses. The premise of giving the diagnostic test to graduate students is that some misconceptions are so rooted in commonsense that even advanced students maintain them. As a result of the high school, undergraduate and graduate physics student interviews, the FCI was modified. Two problems were removed from the FCI due to the majority of students having difficulty with interpretation of the language of the questions.

Data from the second year of field testing, following a pedagogical summer professional development, were combined with data from the first year (Hestenes et al., 1992). Developers did this because a significant difference in student performance was found for only two teachers. Data from a traditional course in Chicago were compared to the Arizona course data, with concluding thoughts that Arizona data were representative of other areas. The FCI developers used a high variety sample at the cost of making some demographics messy to use. The competence ranking system used in FCI development

ranked teachers subjectively on their background and teaching experience. For instance, all but 16 Advanced Placement (AP) students came from teachers ranked in the upper 50% of competence and at the same time in the lower two-fifths socioeconomically. This example illustrates how the variety in the field testing sample complicated the use of demographics to draw conclusions.

Significance of Prior FCI Studies

Deficits in Instruction

Because of the high predictive validity of the FCI, deficits in instruction, rather than student deficits, were likely the culprit in low FCI performance (Hestenes & Halloun, 1995). Teaching method may have served to mediate the effect of gender on physics performance (Else-Quest et al., 2010). Lorenzo, et al. (2006) found that not only did interactive engagement methods increase performance of all students, but the posttest difference of males and females was reduced through interactive engagement. More dramatic was their finding that the posttest differences between males and females was dependent on the degree to which interactive engagement was employed (Lorenzo et al., 2006). Contrary to the findings of Lorenzo et al. (2006), Pollock et al. (2007) found that interactive engagement was not related to smaller differences in FCI performance between males and females. These studies informed the current study in that contradictory evidence for the effect of pedagogy on performance of the FCI warranted a look at the differences for students at TTU.

Context of Culture

Of the student characteristics looked at in a study of physics students at a Canadian university, gender was most related to performance on the FCI pretest (Noack et al., 2009). The authors looked at coursework, grades, and demographics to determine predictors of performance. Completing an upper level physics course in high school explained the greatest portion of the variance in pretest performance between males and females compared to other educational characteristics that were studied. However, the educational characteristics and demographics included in the study were not related to performance gains on the FCI. A curious result of this study was that being born outside of Canada, the setting for the study, was related to low scores on the pretest and low performance gains on the FCI. The authors warned introductory physics instructors about "substantial limitations in terms of the goals that first-year physics instructors can expect" (p. 1274). This study pointed to the culture of the learning environment, learning in an environment different from the one in which you were raised. This finding supported the idea that, if the culture of being female was foreign to the culture of being a physics student, culture in the discipline was related to performance of females.

McCullough (2002) found that math background and performance on the FCI were not related. There was also no interaction between math and gender. Her findings suggested that mathematics background and mathematics performance were not related to the gender differences found on the FCI. This was important for future research, as mathematics ability has long been positively associated with physics stereotypes and negatively associated with being female. Though the issue of female aptitude in

mathematics was put to bed as a myth by the mathematics education research community (Hyde et al., 1990; Hyde et al., 2008), the popularity of the stereotype gave it longevity. The current study sought to address the enduring stereotype of low performance of females in physics by exposing the instructional methods and assessment choices that helped create a perceived difference.

Blue and Heller's (2003) study using matched sampling illuminated the need for novel statistical treatments of pre-post data. By creating male-female pairings based on three pretests, educational background, and locus of control over grades, differences on the posttest were measured. They concluded that when differences between people were controlled for, males and females could learn the same amount of physics. They also concluded that the differences in performance of males and females were due to the differences in gender in the culture rather than differences in being physically male or female. Blue and Heller's (2003) study was relevant to the current study in that the matched sampling design was conducted by controlling for as many background variables as were thought to be relevant.

Gender Force Concept Inventory

The FCI has been a source of gender-bias, sometimes called *gender gap*, studies for a considerable amount of time (Blue & Heller, 2003; Coletta, 2015; Coletta, Phillips, & Steinert, 2012; Dancy, 2004; Dietz, Pearson, Semak, & Willis, 2012; Kost-Smith, 2011; Lorenzo et al., 2006; McCullough, 2002, 2004; McCullough & Foster, 2001; Noack et al., 2009). The FCI was written by males and mostly used males in the images

accompanying questions (Hestenes et al., 1992; McCullough, 2002). McCullough (2001) developed a physics assessment, the GFCI, by altering the gender context of each question on the FCI to a female-oriented scenario while not altering the concept or physics context. The purpose was to alter question scenarios from a formal classroom context to a less formal context one might encounter in the daily routines of life (McCullough, 2001).

Gender Force Concept Inventory Development

The GFCI was created by altering pictorial representations and question scenarios from the FCI to a female-oriented version (McCullough & Foster, 2001). These adjustments to stereotypical female scenarios were intended to address concerns that the contexts of the FCI questions were not concretized to the typical female experience (McCullough, 2004). An illustrative example was that for a cannonball being fired by a man off of a cliff, the context was changed to a bowl being pushed by a baby girl off of a highchair (McCullough, 2001). Rather than an image depicting a ball in a channel found on the FCI, the question wording and image were altered to represent a waterslide (McCullough & Foster, 2001). The FCI depicted a ball falling, while the same question on the GFCI depicted a teddy bear falling (McCullough & Foster, 2001). Cooking, jewelry, and female ice skaters were other female-oriented contexts that were substituted into the physics context (McCullough & Foster, 2001). Development began in 1997 and continued through 2000 (L. McCullough, personal communication, July 30, 2014).

Field Testing of the GFCI

Field testing was conducted at the test developer's university (McCullough, 2004). Further cognizant of the gender threat associated with being in a physics class, she conducted her study in general education classes. There was no mention of the issue of gender during testing. These measures controlled for the male context of the classroom which is typical in physics. The FCI and GFCI were distributed to students without mention of varying test version. Demographic questions were located on the back of the answer sheet to prevent prompting of any feelings of gender threat.

In the 2004 field tests of the GFCI with the FCI, several notable findings warranted further study. Compared to the FCI, females performed better on 13 of 30 questions on the female-oriented version. Males performed better on 5 of 30 questions on the female-oriented version as compared to their performance on the FCI. In summary, females performed the same on the FCI and the GFCI, while performance of males decreased when assessed with the GFCI. When performance was compared at the question level, all possible outcomes occurred. The context of assessment questions and problems used to teach physics were related to differences in performance on some physics concepts (McCullough, 2004). For performance on conceptual physics problems, deep-seated misconceptions were indicated by extremely low scores regardless of gender or test version (McCullough, 2011). Gender interactions were not consistent for different populations (McCullough, 2004). This indicated that the GFCI needed to be studied with other populations, particularly large populations, so that more clear inferences could be drawn.

Summary

This study was situated in the critical discourse of Chambers (2009), Gutiérrez (2008), Ladson-Billings (2006), and Lather (2012) in that achievement gap focus was a deficit model that further promoted stereotypes, placed the burden of equity on the marginalized, and allowed for accumulation of inequity. To support the idea that gaps were simply an accumulation of generational inequity, this research focused on perceived gaps and the classroom practices that mitigate them. Those accumulated inequities were evidenced by the purpose of the first organized public education of females (Tozer et al., 2002) as well as the current culture of college females (Eisenhart & Holland, 1990). STEM pedagogies, good or bad, have helped create the culture of STEM learning that has given learners ideas about culture in the STEM workplace (Shulman, 2005). The way STEM learning was measured influenced our interpretation of student performance. A goal of this study was to look at deficits through the lens of the FCI and GFCI, while considering possible mediators (pedagogy) and moderators (gender) of performance. In short, this study sought to determine if the reported gap was 'real' or simply an outcome of inequitable assessment instruments or deficit thinking.

CHAPTER 3

METHODOLOGY

Introduction

The positivist theoretical framework used in the quantitative study described here was influenced by the evolution of scientific objectivity, outside pressure to solve societal problems, and competitive norms of research funding (Bower, 1998; Lather, 2004b, 2005; St. Pierre, 2002). The methodologies proposed for this *ex post facto* study were selected as a means for knowing more completely how gender contexts and teaching pedagogies possibly influenced performance. The methods described here were used to describe the population of introductory algebra-based physics students and to answer important questions about the differences in physics performance between males and females.

This chapter begins with an introduction to the research goals that guided the methodological choices. The design of the study is then described by painting a picture of the LEAP and traditional classrooms as well as a description of the sampling procedures used to create the Fall 2008–Spring 2014 sample used for analysis. A review of the development, validity, and reliability of the Force Concept Inventory (FCI) and the Gender Force Concept Inventory (GFCI) is presented. Pre-analysis methods are distinguished from other analysis to remind readers that the purpose was to attempt to equate the two groups rather than compare them by background and demographics. An account of the statistical tests used to analyze the data is then presented in the form of a

table rather than narrative in order to clearly convey how the variables were used to answer each research question.

Research Design

This *ex post facto* study used a causal comparative design. If the direction and level of relationship between an independent variable and a dependent variable was affected by the moderator, the moderating variable defined the conditions necessary for the independent variable to operate (Baron & Kenny, 1986). This study determined if gender was a moderating variable, thus defining the conditions necessary for the teaching pedagogy to operate.

Context of the Study

Tennessee Technological University (TTU) is a public four-year, state-funded, comprehensive university located in the southeastern United States, Upper Cumberland region of Tennessee. At the time of this study, the physics department at TTU offered a Bachelor of Science in physics and did not offer graduate studies. The department was engaged in over two decades of pedagogical and curricular action research. Students enrolled in the first semester of an introductory algebra-based physics course sequence for non-majors (PHYS2010) at TTU between Fall 2008 and Spring 2014 were included in this study. There were no math prerequisites for the PHYS2010 course. Students in the analysis sample represented all undergraduate levels and 31 programs at TTU. Students

were enrolled in one of two types of PHYS2010 courses, LEAP or traditional. The traditional and nontraditional LEAP groups were formed through student selection during registration.

All lecture and laboratory portions were called "Algebra-based Physics I" regardless of the type of pedagogy being used. The traditional lecture students participated in a standard lecture section which met for 55 minutes three times a week with a separate 3-hour laboratory once a week. The lecture and the lab were not synchronized. Students who chose a traditional lecture had multiple traditional laboratory sections to choose from when registering. Traditional lecture sections varied in size with a registration cap range of 40–70, while traditional laboratory space allowed for a maximum of 32 students. Due to the lack of seat-space for activity materials and adequate blocks of time to incorporate interaction and discourse into the lecture, traditional sections offered in the lecture hall were prohibitive to the tenants of the LEAP curriculum.

Those who chose a LEAP course registered for a separate laboratory section as well, but the two portions of the course were scheduled so that the laboratory followed directly after the lecture. LEAP lecture sections were noted in the registration system with the following restriction: "Must be taken with Physics 2010-101." This meant that LEAP students took lecture and lab in a continuous block of time, whereas traditional laboratories were conducted independently of the lecture. LEAP lecture and laboratory sections had a maximum enrollment of 40 students, organized in cohorts when multiple sections were offered in a semester. The LEAP sections met for two hours three times per week in an integrated lecture/lab setting with minimal lecturing. The LEAP curriculum

incorporated the tenants of PET, which have been shown to improve conceptual understanding and connection of familiar phenomenon with classroom content (Goldberg et al., 2010). Students participated in small groups working through guided-inquiry activities where they interpreted the data that they collected to develop the physics ideas that formed the learning objectives for the course.

Population & Sample

The study sample included students who completed PHYS2010 during Fall 2008–Spring 2014, which represented the availability of electronic data for student demographics received from the Office of Enrollment Management. Of the PHYS2010 course sections offered in the Fall 2008–Spring 2014 semesters, there were 1-2 sections identified as LEAP and a maximum of one section of traditional. Students repeating the course (n = 5) were included only once, with the second course attempt deleted from the analysis. Cases from the Spring 2009 semester did not have a choice of sections as only LEAP was offered. These cases were excluded from the analysis since the purpose of the study was to establish the probability of choosing LEAP and to adjust for that probability when determining differences on pedagogy and gender. A choice of LEAP or traditional course sections was offered in all other semesters. For the Fall 2008–Spring 2013 courses, 748 students completed both the FCI pretest and posttest. Of that sample, 348 were LEAP students and 400 were enrolled in the traditional section. For the Spring 2014 courses which took the GFCI rather than the FCI, 28 were LEAP students and 45 were enrolled in the traditional section.

Pretests were unannounced and administered during the first class. Posttests were also unannounced and administered at the end of the semester. Students who did not take the pretest or posttest due to absence on testing dates were excluded from the analysis. The rationale behind eliminating these missing cases was that this study was conducted with a pretest posttest factorial design. There were instances where students were missing test data due to oversights during testing that resulted in student scantrons with no identifying codes. This prevented matching of pretests with posttests. These cases were excluded from analysis. Spring 2014 students were given the GFCI pretest and posttest, so analysis of matters pertaining to alternatives to the gender-biased FCI involved those 73 students.

Student background data. Student background variables not available from class rosters were obtained from the Office of Enrollment Management. The following variables were obtained from the Office of Enrollment Management: gender, ethnicity/race, high school GPA, high school name, city and state of high school, highest ACT mathematics score, highest ACT science reasoning score, highest ACT composite score, city and state of permanent residence. Math courses completed from Fall 2008 through Spring 2013 as well as the grade earned in each course were also collected for each student. Any transfer math courses, grade earned, institution of transfer course, and course equivalent awarded by TTU were also obtained. However, these mathematics courses and grades variables were not used in the analysis because of too much missing data and lack of equivalence in the courses taken by different students and at different levels.

Assessment data. First semester introductory algebra-based physics classes from Fall 2008 to Fall 2013 took the FCI as a pretest during the first lecture class of the semester. The FCI was also administered during the last week of classes as a posttest. First semester introductory algebra-based physics classes of Spring 2014 took the GFCI as a pretest during the first laboratory session of the semester. The GFCI was also administered during the last week of classes as a posttest. Laboratory course instructors were involved in the distribution and collection of testing materials during some FCI and GFCI testing sessions. All FCI and GFCI scantrons were maintained by the physics department. The following variables were obtained from the classroom instructor and provided by a physics education researcher at TTU: FCI or GFCI pretest score (*pretest*), FCI or GFCI posttest score (*posttest*), and student responses on FCI or GFCI questions 1–30. The following variables were obtained from class rosters: university ID (*TNumber),* course section and semester (*classID*), pedagogy (*pedagogy*), classification at time of pretest (*ClassLevel*), major at time of pretest (*MajorAtPretest*).

Instruments

Force Concept Inventory Validity & Reliability

The FCI consisted of 30 conceptual questions related to the concept of force. Hestenes et al. (1992) classified the questions into "six conceptual dimensions," (p. 142) which included kinematics, Newton's First Law, Newton's Second Law, Newton's Third

Law, superposition principle, and kinds of force. Multiple questions addressed each dimension, and some questions probed more than one dimension.

Criterion-referenced validity. Half of the 30 FCI questions were taken from the Mechanics Diagnostic (MD), as the purpose of developing the FCI was to improve upon the MD (Hestenes et al., 1992). Pretest and posttest scores from both the FCI and MD were similar enough for developers to establish concurrent criterion-referenced validity. Developers decided that the well-established validity and reliability of the MD warranted no need to reproduce the same procedures for the FCI since the tests were so similar for many students. The MD had high reliability, large KR-20 of 0.8–0.9, according to meta-analysis (Hake, 2007). Students were also given the Mechanics Baseline (MB), another assessment of force dimensions, as a posttest to establish concurrent criterion-referenced validity for the FCI. Others also compared the FCI to the MB test and found a strong positive correlation (r = .91), thus establishing the concurrent criterion-referenced validity of the FCI (Hake, 1998).

Wang and Bao (2010) evaluated FCI performance for over 2,500 calculus-based introductory mechanics students using item response theory. To satisfy the unidimensionality assumption of item response theory, each question was treated as one dimension independent of the other questions. Correlation matrices and eigenvalues indicated that performance differences on all questions were attributed to one variable, force. Goodness of fit analysis indicated that all 30 questions of the FCI fit the model and that data could be fit to the item response model. The high predictive criterion-referenced validity was established by determining a linear relationship between student score on the FCI to the student's fixed proficiency level as determined by the item response model fits

(R^2=0.994). This means that a small group's raw score put into the proficiency equation would yield a proficiency score that could be used with each question-metric for comparison to the norm per question.

Predictive criterion-referenced validity was somewhat established by the algebraic test used to describe student groups during test development (Hestenes et al., 1992). The MB was not correlated to the algebraic test, except for one class of students. For this reason, developers concluded that math ability was unrelated to FCI performance.

Content validity. Prior to field testing that was reported by Hestenes et al. in 1992, early versions of the FCI were written and revised through the use of extensive interviews to determine the most common misconceptions held by students so that those common misconceptions could be used as distractors (Halloun & Hestenes, 1985a; Halloun & Hestenes, 1985b). During field testing, FCI developers interviewed 20 students about their responses to FCI questions (Hestenes et al., 1992). For questions that were answered incorrectly, not a Newtonian response, students rarely explained their answer with Newtonian reasoning. This provided content validity in that non-Newtonian thinking resulted in a non-Newtonian answer choice. For questions that were answered correctly, a Newtonian response, it was "fairly common" for a student to explain their answer with non-Newtonian reasoning (p. 148). This led developers to describe FCI test results "as an upper bound on a student's Newtonian understanding" which had also been established as true for the MD (p. 148). The interviews also helped establish that distractors for questions were working as distractors. The potential for the majority of the questions to differentiate Newtonian from non-Newtonian reasoning established content validity. Interviews confirmed that the questions were being interpreted as planned and

was done to establish content validity of the question wording, which confirmed the questions were testing about the force dimension intended. The FCI did seem capable of distinguishing flawed reasoning from Newtonian reasoning for force concepts, as was the intent when designing this narrowly focused instrument.

Interviews were also conducted with 16 new graduate mechanics students for the purpose of eliciting reasons for their response in order to establish content validity (Hestenes et al., 1992). Students who scored high on the FCI explained their choices within the Newtonian paradigm and were exceedingly successful in the graduate mechanics course. All but two graduate students had buoyancy misconceptions, as was expected based on the FCI developers experiences developing the MB. Three of the graduate students exhibited misconceptions in other dimensions as well as buoyancy. Graduate students with the poorest performance on the FCI were either on academic probation or had failed. The results from interviews illustrated how the FCI discerned Newtonian from Aristotelian thinking. The premise for giving the diagnostic test to graduate students was to establish that some force misconceptions were so rooted in commonsense that little could be done to replace them, even a four-year physics degree. The use of interviews allowed developers to determine that the common dimensional misconceptions were represented as distractors for a test item intended to address that force dimension. This process of assuring that the common misconceptions were included in the choices gave content validity to the test. These students were used to establish content validity by virtue of their extensive coursework in physics. Experienced physics education researchers evaluated the final version for appropriate level and content. This step was important in establishing content validity because the physics education

researchers had extensive exposure to Newtonian conceptual learning studies compared to the group for which the test was designed.

Experimental validity. Harvard physics students were given the FCI and MB on computers for the purpose of determining how much time was necessary to eliminate time as a threat to situational experimental validity. The average computer time was used to inform the minimum time suggestions for each test (Hestenes et al., 1992).

Face validity. Hestenes and Halloun (1995) purported that face validity was established by the critical review of many physics professors, and these reviewers reported no concerns about the appropriate Newtonian responses as item choices.

Construct validity. Each dimension of the overarching concept of force was assessed by multiple-item constructs. In order to establish construct validity, multiple questions described one concept (Hestenes et al, 1998).

Construct analysis of the FCI created much debate, a debate fueled by a difference in opinion regarding the intent of the FCI (Hestenes & Halloun, 1995). Two of the MB test developers who went on to develop the FCI responded to such debate. If a group of Newtonian thinkers, those scoring above 85%, were used to conduct factor analysis of the FCI, developers asserted that there would only be one factor. They felt that factor analysis was only relevant when analyzing within a narrow score range, such as 60–80%, which was considered as somewhat Newtonian (Hestenes & Halloun, 1995). Hestenes and Halloun (1995) also argued that the total FCI score was a better tool than particular constructs or questions for evaluating teaching methods.

Equivalent forms reliability. The FCI was correlated to the Force and Motion Conceptual Evaluation (FMCE) by Thornton, Kuhl, Cummings, and Marx (2009) for the

purpose of establishing equivalent forms reliability of the FCI. A line of best fit for FCI and FMCE posttest data (m = .54) along with a high correlation coefficient (r = .78) indicated that a score on the FCI was a good predictor of a score on a different test of force concepts, the FMCE.

Internal consistency reliability. Since all items on the FCI were scored dichotomously, either correct (1) or incorrect (0), the Kuder-Richardson reliability coefficient, KR-20, could be used to determine the internal consistency of the FCI. The KR-20 value was an estimation of all possible split-halves correlations. This was preferable when it was possible that halves devised in different manners would result in differing correlation values. In order to compare scores on the FCI, KR-20 values should have exceeded 0.80 indicating high internal reliability. KR-20 was preferred because coefficients tend to be lower than for other internal consistency methods, thus improved confidence in the significant coefficients (Engelhardt, 2009). Reliability was established with high Kuder-Richardson coefficients between 0.8–0.9 during the development of the test (Halloun & Hestenes, 1985a). In a later study the test (r = .900), retest (r = .812), and combined test-retest (.865) KR-20 reliability coefficients indicated high internal consistency for the FCI (Lasry, Rosenfield, Dedic, Dahan, & Reshef, 2011).

Test-retest reliability. Test-retest reliability was established by administering the FCI twice to three cohorts of students (N = 110) enrolled in an electricity and magnetism course (Lasry et al., 2011). These students received no additional instruction in the force concepts found on the FCI. The correlation was strong (r = .89), indicating that the FCI was a highly reliable measure of force concepts.

Gender Force Concept Inventory Validity & Reliability

The GFCI consisted of 30 conceptual questions related to the concept of force. The GFCI was created by altering the scenarios and images of the 30 FCI questions, while maintaining the dimension of the force concept being probed (McCullough & Foster, 2001). This meant that the six dimensions of the force concept were thought to be probed by the same questions on the FCI and GFCI, though by different contexts.

Criterion-referenced validity. Question contexts were changed to stereotypical female scenarios without altering the physics content of the problem (McCullough & Foster, 2001).

Content validity. Experts in physics education evaluated the female-oriented contexts to assure that the physics content had not been altered from the original version of the FCI.

Experimental validity. McCullough (2004) conducted her study comparing the GFCI and FCI in general education classes rather than in physics classrooms. The FCI and GFCI were distributed to students in ABAB order, with no mention of test version differences. Fifteen minutes into the test, students were asked to answer demographic questions after the test questions. These measures were put into place to attempt to control for gender threat and other stereotype threats that may be present in a physics testing environment for some groups of students.

Face validity. Contexts were as stereotypically female as was conceivable to allow for the largest observable shift due to context (McCullough, 2002).

Reliability. Cronbach's alpha values showed larger correlations on the GFCI compared to the original version of the FCI.

Methods

Pre-Analysis Data Screening

Variable transformations. Categorical variables (timespan, pedagogy, classification at time of pretest, major, gender, race, rurality) were transformed by creating dummy codes. Shapiro-Wilk's test, and histograms were used to evaluate normality of continuous variables in each analysis (Razali & Wah, 2011). Skewness and kurtosis were also used to evaluate normality. Missing cases were deleted from analysis when cases comprised less than 5% of the variable in question. When missing cases exceeded 5% of cases within a continuous predictor, these cases were replaced by the mean.

LEAP Program Evolution. There was concern that evaluating for effects of the LEAP program was confounded by the data-driven evolution of the LEAP curriculum and instruction. To account for this, course sections were binned (grouped) by timespan in order to test for significant changes in the LEAP curriculum. Since the curriculum changed in Fall of 2010, the decision was to look into whether or not timespan was related to predicting the likelihood of a student choosing LEAP over traditional sections. Logistic regression results indicated that timespan was a significant factor in predicting if a student would choose a LEAP section. Data from Fall 2010 and onward were used to answer the research questions of this study.

Propensity scoring. In the current study, PHYS2010 students self-selected a course while transparency of the differences between course sections did not exist. It is important to acknowledge and attempt to control for bias due to selection of course sections (Clouston, 2013). Treatment effects in an observational study cannot be estimated bias-free by direct comparison of groups on the outcome measure of FCI performance (Austin, 2011a). The purpose of propensity scoring in this study was to create one covariate from a larger set of covariates. For this study, it was assumed that characteristics may have existed that predisposed a student to select either a LEAP or a traditional section. If a statistically significant difference was found between LEAP and traditional courses while controlling for student covariates, a portion of FCI performance differences may have been attributed to pedagogy (Austin, 2011a; Rudner & Peyton, 2006).

Using logistic regression, a propensity score was calculated from the covariates through prediction of a dichotomous dependent variable (pedagogy). A probability of registering for either section was generated, with the probability of registering for LEAP being the propensity score. Assumptions of propensity scoring methods were similar to the assumptions of regressing outcomes on treatment conditions and confounders: treatment assignment was ignorable due to the inclusion of all known confounders in the model (Austin, 2011a). The process of propensity scoring did not reduce bias, and the covariates used to generate the score were no more appropriate than before the analysis (Pearl, 2009). The covariates used for propensity scoring analysis included: Timespan (*Timespan)*; semester and year of pretest (*Semester*), classification at time of pretest (*Level*), division at time of pretest (*Division*), race (*Race*), gender (*Gender*), engineering

major (*Engin*), pre-professional major (*PreProf*), exercise sciences major (*ExSci*), rurality (*rurality*), high school GPA (*HighSchoolGPA*), highest ACT mathematics score (*ACTMath_1*), highest ACT science reasoning score (*ACTScience_1*), highest ACT composite score (Highest*ACTComposite*). The descriptions of each of these variables are given in Chapter 4. For details regarding the generation of a propensity score for each student and steps for evaluating the model generated, the following sources were helpful: Austin (2011a, b), Clouston (2013), Rosenbaum & Rubin (1983a, b). Rubin (2007), Rudner & Peyton (2006), and Wuensch (2014).

As an alternative to propensity score matching, the propensity score was used to look at the effects of gender and pedagogy on FCI and GFCI performance while blocking on the propensity score. By using the propensity score as one covariate that described a large number of relevant, feasible covariates, the loss of statistical power that follows loss of degrees of freedom was avoided (Austin, 2011a). Predicted group membership (*PGR_3)*, which was the propensity score binned at a cut point of .50, was utilized when interactions between the propensity score and predictors warranted further analysis. This procedure was followed for FCI and GFCI data independently and described in detail in the following sections.

Data Analysis

Table 1 shows the analyses that were conducted to answer each research question.

Table 1
Summary of Analysis Methods Used to Answer Research Questions

Purpose	Analysis Technique	IVs	DV
Research Question 1* *Is there a difference in the performance of males and females on the FCI?* Research Question 3 *Is there a difference between students taught using LEAP pedagogy versus a traditional pedagogy on the FCI?* Research Question 4 *Is the difference in performance on the FCI between students who were in the LEAP versus traditional the same for males and females?*	Three-way analysis of variance	Pedagogy Gender Predicted group membership	FCI normalized gain
	Four-way analysis of variance	Pedagogy Gender Predicted group membership Pretest binned	FCI posttest
Research Question 2* *Is there a difference in the performance of males and females on the GFCI?* Research Question 5 *Is there a difference between students taught using LEAP pedagogy versus a traditional pedagogy on the GFCI?* Research Question 6 *Is the difference in performance on the GFCI between students who were in the LEAP versus traditional the same for males and females?*	Three-way analysis of variance	Pedagogy Gender Predicted group membership	GFCI normalized gain
	Four-way analysis of variance	Pedagogy Gender Predicted group membership Pretest binned	GFCI posttest
Research Question 7 *Is there a difference between males and females on particular constructs of the FCI?* Research Question 8 *Is there a difference between students taught using LEAP pedagogy versus a traditional pedagogy on particular constructs of the FCI?*	Principal component factor analysis	FCI Items 1–30	
	Independent samples *t*-test	Gender	FCI factors 1–8
	Independent samples *t*-test	Pedagogy	FCI factors 1–8

*Analysis for question 1 also answered questions 3 and 4. Analysis for question 2 also answered questions 5 and 6.

Summary of Methods

Through a descriptive *ex post facto* design for the analysis of pedagogy and gender differences, performance of males and females was compared. Differences could not be described without attention to variables which confound the analysis. For that reason, propensity scoring methods were used to control for the probability of students' self-selection into the LEAP sections of the physics course. The intent was to describe performance as a function of pedagogy, rather than as a function of deficits. Outside of the deficit model, the blame for the underperformance of females was attributed to discourses of power as well as less-than-critical ways of evaluating learning and schooling.

In Chapter 4, an analysis of data is presented. Chapter 5 begins with an alternative representation that foreshadows the findings and discussion presented.

CHAPTER 4

DATA ANALYSIS

Pre-Analysis Data Screening

Variable Descriptions

Timespan. *ClassID* was used to create the dummy coded *Timespan* variable (Fall2008–Spring2010 = 1, Fall2010–Spring2014 = 2). The binning cut point was based on ongoing action research which led to significant changes in LEAP curriculum beginning Fall 2010.

Pedagogy. *ClassID* was used to create *Pedagogy,* a dummy coded dichotomous variable for traditional (0) and LEAP (1) sections of PHYS2010.

Class level. Classification at the time of pretest, *ClassLevel*, was used to create a dummy coded variable, *Level*. The dichotomous *Division* variable was created from *Level* by dividing students into lower (0) and upper (1) divisions. *Level* was preferred for analysis because it was more disaggregated.

Race. Self-reported race descriptions and corresponding university race codes were used to create the dichotomous dummy coded variable, *Race*. Students reporting white as their only race were coded as White (0). Students reporting races other than white, or multiple races including white, were coded as Other Race (1). Missing cases

with *ethnicity description* of "not Hispanic Latino" were treated as missing cases while "white, non Hispanic" were coded as White. Other Race made up 10.6% of the sample.

Gender. A dummy coded *Gender* variable was create for female (0) and male (1).

Major. *MajorAtPretest* originally had 37 levels representing 37 programs. The variable was then used to create *MajorAtPretest2* (Chemical Engineering, Engineering Technology, and Mechanical Engineering = 1; Pre-Medicine, Pre-Pharmacy, Pre-Optometry, Nursing, General Health Studies, Biology, Chemistry, Agriculture, Pre-Dentistry, Physics = 2; ExerciseSci PhysEd Wellness, Pre-Occupational Therapy, Pre-Physical Therapy = 3; all other majors = 4). Organization of groupings was informed by the stratification of groups experienced by curriculum developers, particularly the career goal motivations for enrolling in a physics course. For example, physics was not required for nursing or agriculture degrees but was necessary for Physician's Assistant and Veterinary Medicine programs. Pure chemistry majors took calculus-based physics, so any chemistry majors enrolled in PHYS2010 were on an applied chemistry path. These students were identified as pre-professional because of where they were going after the undergraduate degree. *MajorsAtPretest2* was transformed to *Major* by dummy coding those categories (Engineering = 1, Pre-Professional = 2, Exercise Sciences = 3, Other Majors = 4). Categorization of the 31 programs, representing the students composing the final sample, into those four groups is detailed in Appendix A. Effect coding was then used to create three new variables: *Engin*, *PreProf*, *ExSci*.

Rurality. The high school name, city, and state were used to determine the urban/rural classification using the National Center for Education Statistics (NCES) online tool. Private schools not found in the NCES database were assigned the same code

as the school on the same block or street. Homeschools were coded as rural since the reason for investigating this variable was to probe the possibility that the school and community size influenced the choice of a large (traditional) or small (LEAP) section of PHYS2010. Those with GED Equivalent listed as the high school were coded as urban based on the school corresponding to their city and county of permanent address. If more than one high school was attended, the case was assigned the classification of the first school listed. Some classifications troubled common sense. Regardless of the ways one might define rurality, the NCES is consistent across schools. A dummy coded *Rurality* variable was created for urban (0) and rural (1).

High school GPA. High schools calculated and reported GPA to TTU. Any GPA above 4.0 was truncated for university records. Four cases remained above 4.0 and were wildly impossible given any means for calculating GPA, such as 59.6 and 525. These were treated as missing cases.

Mathematics courses. All mathematics courses taken from Fall 2008 to Spring 2013 and corresponding course grades for each student were used to create two new variables, *GradeinHighest1000LevelMathCoursePriorPHYS2010* and *GradeinHighest2000LevelMathCoursePriorPHYS2010*. An increase in course number indicated a more advanced mathematics course. Transfer, Advanced Placement, and International Baccalaureate credits were included in the determination. All mathematics courses taken during or after PHYS2010 were eliminated from the determination for each student. This method did not allow for distinguishing between an A in College Algebra and an A in Calculus II, though easily distinguishable measures of mathematics ability. Mathematics course data prior to Fall 2008 were not available, which further questioned

the reliability of this variable. Due to how this could have confounded the analysis, this variable was not included in the analysis pending looking into *ACTMath* as a measure of mathematics performance.

FCI and GFCI questions. Students selecting two choices for one question were assigned one of their choices. The first case with multiple selections was assigned their first answer choice. The second case with multiple selections was assigned their second answer choice. This alternation continued until all cases had only one answer per question. Letters of choice for *Q1* through *Q30* were dummy coded as correct (1) and incorrect (0) to form *Item1* through *Item30*.

Missing Cases

Missing cases for *Race* (n = 3, 0.5%), *Rurality* (n = 9, 1.5%), and *HighSchoolGPA* (n = 8, 1.3%) were deleted from analysis using those variables since those cases made up less than 5% of cases. Missing cases for *ACTMath* (n = 32, 5.3%) were greater than 5% of cases and were replaced by the mean of the variable, creating *ACTMath_1*.

Propensity Scoring

The purpose of propensity scoring in this study was to create one covariate from a larger set of covariates. The covariates considered for propensity scoring analysis

included: Timespan (*Timespan)*; semester and year of pretest (*Semester*), classification at time of pretest (*Level*), division at time of pretest (*Division*), race (*Race*), gender (*Gender*), engineering major (*Engin*), pre-professional major (*PreProf*), exercise sciences major (*ExSci*), rurality (*Rurality*), high school GPA (*HighSchoolGPA*), highest ACT mathematics score (*ACTMath_1*), highest ACT science reasoning score (*ACTScience_1*), and highest ACT composite score (Highest*ACTComposite*). Logistic regression was conducted to determine the accuracy of the independent variables predicting that a student would choose a LEAP section.

A preliminary multiple regression was used to calculate Mahalanobis' distance and examine multicollinearity of the four continuous predictors. Tolerance values indicated that multicollinearity was not problematic. Five cases with Mahalanobis distances that exceeded the critical chi-squared, $\chi^2_{crit} = 18.467$, $df = 4$, $p < .001$, were identified as multivariate outliers. These cases were not excluded from the analysis. The Box-Tidwell test was used to test the assumption of linearity of the log odds and the continuous predictor variables. Logistic regression indicated that the assumption of linearity was met for each continuous variable and its natural log: high school GPA [$B = -.903$, Wald = .256, $df = 1$, $p = .613$, $e^B = .405$], ACT math [$B = .003$, Wald = .000, $df = 1$, $p = .989$, $e^B = 1.003$], ACT science [$B = -.307$, Wald = .278, $df = 1$, $p = .270$, $e^B = .736$], and highest ACT composite [$B = .389$, Wald = .342, $df = 1$, $p = .256$, $e^B = 1.475$]. Previous pre-analysis data screening merited follow-up regarding the likelihood that student selection of LEAP over traditional was influenced by the timespan in which the selection was made. Logistic regression indicated that timespan [$B = .663$, Wald = 4.409, $df = 1$, $p = .036$, $e^B = 1.941$] significantly increased the chance of choosing LEAP

membership by nearly double. After Spring 2010, students were 1.941 times more likely to choose LEAP over traditional sections than students enrolling prior to the curriculum change. This informed the decision to omit Fall 2008–Spring 2010 cases from further analysis.

Histograms indicated that high school GPA was negatively skewed, while ACT math and ACT science reasoning scores are normally distributed. Missing cases comprising less than 5% of Fall 2010–Spring 2014 cases were deleted listwise. The Hosmer and Lemeshow test Chi-square was significant, $\chi^2 = 9.444$, $df = 8$, $p = .306$, indicating good fit of data with the linear model. Though model fit statistics were large and indicative of a poor-fitting model (-2 Log Likelihood = 765.342, $\chi^2(9) = 41.425$, $p < .001$), the generated model was statistically reliable in predicting those who chose LEAP and those who did not choose LEAP.

Regression coefficients are presented in Table 2. Wald statistics indicated that two variables, *Race* and *Rurality*, significantly predicted group membership. Though Wald statistics indicated that students from rural high schools were nearly twice as likely ($e^B = .57$, where Rural = 1 and Traditional = 0) to choose traditional PHYS2010 sections as students from urban high schools, the NCES classification of some feeder counties should cause caution in interpreting the odds ratio. Students reporting races other than only white were almost twice as likely ($e^B = 1.97$, where Other Race = 1 and LEAP = 1) to choose LEAP, however the 90–10 (white-other than white) split of the variable begs caution in interpreting these statistics.

Table 2
Logistic Regression Coefficients

	B	Wald	*df*	*p*	*OR*
Level	.15	1.97	1	.160	1.17
Race	.68	4.86	1	.028	1.97
Gender	-.28	1.96	1	.162	.76
Engin	-.37	2.80	1	.094	.69
PreProf	.14	.89	1	.345	1.15
ExSci	.27	2.32	1	.127	1.31
Rurality	-.57	9.89	1	.002	.57
HighSchoolGPA	.41	2.97	1	.085	1.51
ACTMath_1	-.06	3.57	1	.059	.94
ACTScience_1	.01	.06	1	.802	1.01
Constant	-.48	.25	1	.619	.62

Few meaningful predictors could be gleaned from the model. However, this was not much of a concern as the overall goal of this step was to come up with a conglomerate, the propensity score of group membership. Table 3 shows that the model could not predict membership in LEAP without these variables any better than a penny toss, as indicated by 52.5% overall predictability. Table 4 shows that when including the predictors in the model, 63.0% of cases were correctly classified. This indicated that when these covariates were known, there was a 10.5% increase in being able to correctly predict that a student would choose LEAP rather than traditional. Albeit the individual variables may not be significant, collectively they contributed to the predictive capability of group membership. This study was not a randomized controlled trial since students self-selected a group; controlling for the propensity score was an attempt to equate these groups after that fact.

Table 3
Logistic Regression Classification Table with No Predictors

	Predicted Pedagogy (Block 0)		
Observed Pedagogy	Traditional	LEAP	% Correct
Traditional	306	0	100.0
LEAP	277	0	0
Overall Percentage			52.5

Table 4
Logistic Regression Classification Table with Predictors

	Predicted Pedagogy (Block 1)		
Observed Pedagogy	Traditional	LEAP	% Correct
Traditional	214	92	69.9
LEAP	124	153	55.2
Overall Percentage			63.0

A dichotomous group membership variable, *PGR*_3, and a propensity score were generated for 583 cases. Inspection of the histogram as well as skewness and kurtosis values indicated normality of the propensity score. Propensity score matching for this timespan of students was not appropriate since the small explained variability caused difficulty in creating good nearest neighbor matches. Setting the caliper at .2 times the standard deviation of the logit, the optimal caliper for matching students (Austin, 2011b), would have matched a student to a large portion of the other students. The propensity score was then investigated as a covariate. Due to covariate interaction with the independent variables of pedagogy and gender, the dichotomous predicted group membership (predicted to choose LEAP versus predicted to choose traditional) was then used as an independent variable for both FCI and GFCI data analysis.

Results

Force Concept Inventory Analysis

The sample for FCI analysis included students from LEAP (N = 258) and traditional (N = 272) course sections, with males (N = 256) and females (N = 274) representing roughly equal portions of the overall sample. Performance means for traditional and LEAP sections are presented in Table 5. Table 6 shows those means parsed out by gender with respect to pedagogy.

Table 5
Unadjusted FCI Means by Pedagogy

Pedagogy	*N*	Pretest (*SD*)	Posttest (*SD*)	Normalized Gain (*SD*)
Traditional	272	8.10 (4.05)	13.45 (5.43)	.24 (.23)
LEAP	258	7.57 (3.66)	19.33 (5.25)	.53 (.23)

Table 6
Unadjusted FCI Means by Gender

Pedagogy	Gender	*N*	Pretest (*SD*)	Posttest (*SD*)	Normalized Gain (*SD*)
Traditional	Female	128	6.27 (2.87)	11.66 (4.59)	.21 (.22)
	Male	144	9.73 (4.25)	15.03 (5.63)	.27 (.24)
LEAP	Female	146	6.29 (2.66)	18.40 (5.05)	.51 (.22)
	Male	112	9.24 (4.11)	20.54 (5.29)	.55 (.23)

The initial intent was to use the propensity score variable as a covariate to adjust for the probability of students joining the section of LEAP. However, the assumption of homogeneity of regression slopes was violated, indicating that the propensity score interacted with levels of the independent variables. The propensity score was then stratified, creating predicted group membership (*PGR_3)* at a cut point of .50, due to interactions between the propensity score and predictors warranting further analysis. Predicted group membership, blocked on the propensity scores as either predicted to choose LEAP (.50–1.0) or predicted to choose traditional (.00–.49), was used as an additional independent variable rather than as a continuous covariate.

A three-way factorial ANOVA by gender and pedagogy with blocking on the propensity score was conducted. Table 7 summarizes the main effects and interactions; ANOVA revealed no main effect for gender [$F(1, 506) = .73$, $MS_e = .04$, $p = .395$, partial $\eta^2 = .001$] but did reveal main effects for pedagogy [$F(1, 506) = 204.09$, $MS_e = 10.16$, $p < .001$, partial $\eta^2 = .287$] and predicted group membership [$F(1, 506) = 13.54$, $MS_e = .67$, $p < .001$, partial $\eta^2 = .026$]. An interaction of pedagogy, gender, and predicted group membership, $F(1, 506) = 3.91$, $MS_e = .20$, $p = .048$, partial $\eta^2 = .008$, was slightly significant but explained less than 1% of the differences on normalized gains. Pedagogy explained 28.7% of the difference in FCI normalized gains, while gender accounted for no variance. Predicted group membership accounted for 2.6% of variation in normalized gain means. For students predicted to have chosen a traditional section, those who were taught with LEAP pedagogy had higher normalized gains than those taught with a traditional pedagogy. The increase in normalized gain accounted for by pedagogy was true equally for both genders as confirmed by the lack of interaction [$F(1, 506) = .52$,

MS_e = .03, p = .470, partial η^2 = .001]. For those predicted to enroll in a LEAP section, both males and females had higher normalized gains when taught with LEAP pedagogy.

The assumption of homogeneity of regression was again not met for the analysis to test for differences in FCI performance on the posttest while covarying on the pretest by pedagogy, gender, and PGR_3 [$F(3, 523) = 3.22$, $MS_e = 68.04$, p = .023]. Subsequently, a four-way ANOVA was conducted using binned FCI pretest scores as the fourth variable. The pretest score was stratified to create a dichotomous ordinal variable (*PretestBin)*, at a cut point of seven correct items which represented the natural break nearest the mean (low ≤ 7, high > 7).

Table 7
ANOVA Summary for FCI Normalized Gain Blocked on Propensity Score

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Pedagogy	10.16	1	10.16	204.09	.000	.287
Gender	.04	1	.04	.73	.395	.001
PGR_3	.67	1	.67	13.54	.000	.026
Pedagogy x Gender	.03	1	.03	.52	.470	.001
Pedagogy x PGR_3	.06	1	.06	1.13	.289	.002
Gender x PGR_3	.04	1	.04	.86	.355	.002
Pedagogy x Gender x PGR_3	.20	1	.20	3.91	.048	.008
Error	25.18	506	.05			
Total	112.15	514				

Model R^2 = .319 (R^2_{adj} = .310)

Table 8 summarizes the main effects and interactions. The main effect of gender was not significant [$F(1, 498) = 3.80$, $MS_e = 90.55$, $p = .052$, partial $\eta^2 = .008$] but there were significant main effects for pedagogy [$F(1, 498) = 182.19$, $MS_e = 4344.84$, $p < .001$, partial $\eta^2 = .268$] predicted group membership [$F(1, 498) = 12.55$, $MS_e = 299.34$, $p < .001$, partial $\eta^2 = .025$] and binned pretest [$F(1, 498) = 31.48$, $MS_e = 750.66$, $p < .001$, partial $\eta^2 = .059$]. An interaction of pedagogy, gender, and predicted group membership was again significant, $F(1, 498) = 5.55$, $MS_e = 132.35$, $p = .019$, partial $\eta^2 = .011$, but only explained 1.1% of the differences in posttest means. Pedagogy explained 26.8% of the difference in FCI posttest performance, while gender did not explain performance differences. Performance on the pretest accounted for 5.9% of the variance, and predicted group membership accounted for 2.5% of variation in posttest means. These main effects needed to be interpreted with caution due to the significant interaction, though the interaction accounted for slight amounts of differences due to pedagogy, gender, and predicted group membership. For students predicted to choose a traditional section, those who were taught with LEAP pedagogy had higher posttest scores than those taught with a traditional pedagogy. For those predicted to enroll in a LEAP section, both males and females had higher posttest performance when taught with LEAP pedagogy. The increase in posttest performance accounted for by pedagogy was the same for both genders as confirmed by the lack of interaction between pedagogy and gender [$F(1, 498) = .87$, $MS_e = 20.76$, $p = .351$, partial $\eta^2 = .002$].

Table 8
ANOVA Summary for FCI Posttest Blocked on Propensity Score & Pretest

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Pedagogy	4344.84	1	4344.84	182.19	.000	.268
Gender	90.55	1	90.55	3.80	.052	.008
PretestBin	750.66	1	750.66	31.48	.000	.059
PGR_3	299.34	1	299.34	12.55	.000	.025
Pedagogy x Gender	20.76	1	20.76	.87	.351	.002
Pedagogy x PretestBin	18.01	1	18.01	.76	.385	.002
Pedagogy x PGR_3	19.76	1	19.76	.83	.363	.002
Gender x PretestBin	5.39	1	5.39	.23	.635	.000
Gender x PGR_3	41.17	1	41.17	1.73	.189	.003
PretestBin x PGR_3	89.97	1	89.97	3.77	.053	.008
Pedagogy x Gender x PretestBin	18.72	1	18.72	.79	.376	.002
Pedagogy x Gender x PGR_3	132.35	1	132.35	5.55	.019	.011
Pedagogy x PretestBin x PGR_3	4.89	1	4.89	.21	.651	.000
Gender x PretestBin x PGR_3	.01	1	.01	.00	.987	.000
Pedagogy x Gender x PretestBin x PGR_3	1.28	1	1.28	.05	.817	.000
Error	11876.11	498	23.85			
Total	156211.00	514				

Model R^2 = .380 (R^2_{adj} = .362)

Gender Force Concept Inventory Analysis

GFCI data were analyzed using the same methods as FCI data. Cases with no propensity score (*n* = 4) composed > 5% of the sample, and these were transformed to the sample mean. The sample for GFCI analysis included students from LEAP (*N* = 28) and traditional (*N* = 45) course sections, with males (*N* = 45) and females (*N* = 28) being represented equally in LEAP but not in traditional sections as a whole. Performance

means for traditional and LEAP sections are presented in Table 9. Table 10 shows those means parsed out by gender with respect to pedagogy.

Although the assumption of homogeneity of regression slopes was met for this analysis, for consistency with the other tests, a factorial ANOVA was conducted, with the predicted group membership as the additional independent variable. This variable was utilized in the same manner as was done for FCI analysis.

Table 9
Unadjusted GFCI Means by Pedagogy

Pedagogy	*N*	Pretest (*SD*)	Posttest (*SD*)	Normalized Gain (*SD*)
Traditional	45	7.20 (3.06)	10.24 (3.99)	.12 (.19)
LEAP	28	8.29 (3.23)	18.07 (5.17)	.46 (.21)

Table 10
Unadjusted GFCI Means by Gender

Pedagogy	Gender	*N*	Pretest (*SD*)	Posttest (*SD*)	Normalized Gain (*SD*)
Traditional	Female	13	4.46 (1.90)	8.62 (4.65)	.16 (.18)
	Male	32	8.31 (2.73)	10.91 (3.56)	.10 (.20)
LEAP	Female	15	7.07 (2.28)	16.73 (5.35)	.42 (.24)
	Male	13	9.69 (3.66)	19.62 (4.68)	.50 (.19)

A three-way factorial ANOVA by gender and pedagogy with blocking on the propensity score was conducted. Table 11 summarized the main effects and interactions. ANOVA showed no main effect for gender [$F(1, 61) = .09$, $MS_e = .00$, $p = .760$, partial $\eta^2 = .002$] or predicted group membership [$F(1, 61) = 2.65$, $MS_e = .10$, $p = .108$, partial $\eta^2 = .042$] but did reveal a main effect for pedagogy [$F(1, 61) = 22.28$, $MS_e = .86$, $p < .001$, partial $\eta^2 = .267$]. No interactions of predictors of normalized gains were identified. Pedagogy explained 26.7% of the difference in GFCI normalized gains, while gender and predicted group membership accounted for nothing. The difference in normalized gain accounted for by pedagogy did not change based on gender as confirmed by the lack of interaction between pedagogy and gender [$F(1, 61) = .54$, $MS_e = .02$, $p = .467$, partial $\eta^2 = .009$].

Table 11
ANOVA Summary for GFCI Normalized Gain Blocked on Propensity Score

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Pedagogy	.86	1	.86	22.28	.000	.267
Gender	.00	1	.00	.09	.760	.002
PGR_3	.10	1	.10	2.65	.108	.042
Pedagogy x Gender	.02	1	.02	.54	.467	.009
Pedagogy x PGR_3	.03	1	.03	.81	.372	.013
Gender x PGR_3	1.37E-5	1	1.37E-5	.00	.985	.000
Pedagogy x Gender x PGR_3	.03	1	.03	.85	.361	.014
Error	2.36	61	.04			
Total	8.22	69				

Model $R^2 = .452$ ($R^2_{adj} = .389$)

Although the assumption of homogeneity of regression slopes was met for analysis of GFCI posttest differences, for consistency with the other tests a factorial ANOVA was conducted with predicted group membership as an independent variable. To account for pretest differences, the pretest score was stratified (*PretestBin)* at a cut point of seven and was utilized as an additional dichotomous independent variable. A four-way factorial ANOVA by gender, pedagogy, binned FCI pretest, and with blocking on the propensity score was conducted. Table 12 summarizes the main effects and interactions. The results indicated no main effect for gender [$F(1, 56) = .65$, $MS_e = 10.81$, $p = .422$, partial $\eta^2 = .012$], predicted group membership [$F(1, 56) = 1.44$, $MS_e = 23.89$, $p = .235$, partial $\eta^2 = ..025$], or binned pretest [$F(1, 56) = 1.05$, $MS_e = 17.33$, $p = .311$, partial $\eta^2 = .018$] but did reveal a main effect for pedagogy [$F(1, 56) = 25.92$, $MS_e = 429.03$, $p < .001$, partial $\eta^2 = .316$]. No interactions were identified. Pedagogy explained 31.6% of the difference in GFCI posttest performance, while gender did not explain performance differences. The increase in posttest performance accounted for by pedagogy was the same for both genders as confirmed by the lack of interaction between pedagogy and gender [$F(1, 56) = .00$, $MS_e = .01$, $p = .985$, partial $\eta^2 < .001$]. Performance on the pretest and predicted group membership did not account for variance in GFCI posttest means.

Table 12
ANOVA Summary for GFCI Posttest Blocked on Propensity Score & Pretest

Source	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Pedagogy	429.03	1	429.03	25.92	.000	.316
Gender	10.81	1	10.81	.65	.422	.012
PretestBin	17.33	1	17.33	1.05	.311	.018
PGR_3	23.89	1	23.89	1.44	.235	.025
Pedagogy x Gender	.01	1	.01	.00	.985	.000
Pedagogy x PretestBin	10.26	1	10.26	.62	.434	.011
Pedagogy x PGR_3	15.92	1	15.92	.96	.331	.017
Gender x PretestBin	.90	1	.90	.05	.817	.001
Gender x PGR_3	.01	1	.01	.00	.985	.000
PretestBin x PGR_3	.02	1	.02	.00	.974	.000
Pedagogy x Gender x PretestBin	.00	0	-	-	-	.000
Pedagogy x Gender x PGR_3	20.83	1	20.83	1.26	.267	.022
Pedagogy x PretestBin x PGR_3	.00	0	-	-	-	.000
Gender x PretestBin x PGR_3	.00	0	-	-	-	.000
Pedagogy x Gender x PretestBin x PGR_3	.00	0	-	-	-	.000
Error	927.08	56	16.56			
Total	13692.00	69				

Model R^2 = .560 (R^2_{adj} = .466)

Concept Inventory Factor Analysis

Testing for Differences on Force Concept Inventory Constructs

Principal components factor analysis utilizing equamax rotation was conducted to determine what underlying structure exists for student responses on the 30 multiple choice questions of the FCI. The analysis produced an eight-component solution (Table

13), which was evaluated with the following criteria: eigenvalue, variance, scree plot, and residuals. Eight components had an eigenvalue > 1. Only three communalities (n = 530, df = 30) were ≥ .60, which made the application of the eigenvalue criteria questionable. After rotation, the first component accounted for 7.89% of the total variance in the original variables, while the second and third components accounted for 7.27% and 7.19% respectively. In evaluating the remaining variances, an eight-component solution accounted for 50.4% of the total variances of the original variables. Though the scree plot indicated a slope between factors 8 and 9 which appeared similar to that slope found between factors 7 and 8, the additional components explained less than one original variable, as indicated by eigenvalues, and were not included in further analysis. Though slightly more than 50% of residuals were greater than 0.05, ten or more loadings < |.4| were present for each factors. Factors were named by evaluating the content and common misconceptions associated with the test items that had high loadings, correlation coefficient of absolute value ≥ .400, on each factor.

Table 13 represents loadings for each factor, all of which are positive. Factor 1 consisted of six test items and addressed combinations of both vertical and horizontal forces acting on an object, *Vertical and Horizontal Force Combinations*. Factor 2 consisted of three test items and addressed combinations of forces in one direction and their resulting effect on an object's motion, *One Direction Force Combinations*. Factor 3 consisted of three test items and addressed fundamentals of Newton's Third Law of motion, *Newton's Third Law*. Factor 4 consisted of three test items which did not have an obvious relationship. It is difficult to then suggest why students related these FCI questions together, so this factor was named *Other Situations*. Factor 5 consisted of four

test items and involved interpreting the motion of an object after a force ceases to act, *Force Ceases*. Factor 6 consisted of three test items and involved interpreting strobe diagrams to obtain velocity and acceleration, *Interpreting Strobe Diagrams*. Factor 7 consisted of two test items that addressed the effect of no force or no net force on an object's resulting motion, *No Net Force*. Factor 8 consisted of two test items which addressed the effect of a sideways force on an object moving at a constant speed, *Force Perpendicular to Motion*.

Independent samples *t*-tests were conducted to determine differences in factor scores for gender and pedagogy. Levene's test for equality of variances was used to determine if the assumption of homoscedasticity was met for each factor, with the *t*-test statistic for unequal variances used for any factors with violation of the assumption. The two-tailed *t*-test statistic was evaluated at the .00625 alpha level, 0.05 corrected for eight *t*-tests, to identify if differences in FCI factor means between 1) LEAP and traditional and 2) females and males existed.

Table 13
Force Concept Inventory Factor Loadings

FCI Item	Loading	Factor (initial eigenvalue)	% of variance explained
Item 18	.710	*Vertical and Horizontal Force Combinations* (6.08)	7.89%
Item 5	.627		
Item 13	.554		
Item 30	.544		
Item 11	.491		
Item 10	.455		
Item 25	.703	*One Direction Force Combinations* (1.71)	7.27%
Item 17	.679		
Item 26	.624		
Item 4	.762	*Newton's Third Law* (1.54)	7.19%
Item 15	.754		
Item 28	.691		
Item 3	.654	*Other Situations* (1.31)	6.27%
Item 23	.581		
Item 22	.538		
Item 7	.634	*Force Ceases* (1.24)	5.93%
Item 12	.607		
Item 6	.585		
Item 27	.419		
Item 19	.702	*Interpreting Strobe Diagrams* (1.12)	5.91%
Item 20	.615		
Item 9	.498		
Item 29	.680	*No Net Force* (1.08)	5.25%
Item 24	.583		
Item 21	.676	*Force Perpendicular to Motion* (1.05)	4.71%
Item 14	.491		

Results of independent samples *t*-tests for pedagogy are summarized in Table 14. Differences between students taught with a traditional approach and those taught using LEAP pedagogy were significant on *Vertical and Horizontal Force Combinations* [*t* = -

5.21, $p < .001$], *One Direction Force Combinations* [$t = -9.93$, $p < .001$], *Newton's Third Law* [$t = -6.78$, $p < .001$], and *Other Situations* [$t = -3.47$, $p = .001$], with students in traditional sections scoring significantly lower on those factors. Differences were also significant on *Force Perpendicular to Motion* [$t = 6.40$, $p < .001$], with students in LEAP sections scoring lower than students in traditional sections. The *t*-test did not indicate a statistically significant difference between traditional sections and LEAP sections on *Force Ceases*, *Interpreting Strobe Diagrams*, and *No Net Force*.

Table 14
Pedagogy Differences on FCI Factors

FCI Factors	Traditional	LEAP	*t*-test	*p*
Vertical and Horizontal Force Combinations	-.215	.227	-5.21	.000
One Direction Force Combinations	-.388	.409	-9.93	.000
Newton's Third Law	-.273	.288	-6.78	.000
Other Situations	-.145	.153	-3.47	.001
Force Ceases	-.040	.042	-.94	.347
Interpreting Strobe Diagrams	-.079	.083	-1.87	.062
No Net Force	-.095	.100	-2.26	.024
Force Perpendicular to Motion	.260	-.274	6.40	.000

Results of independent samples *t*-tests for gender are summarized in Table 15. Differences between males and females were significant on *Other Situations* [$t = -4.85$, $p < .001$], *Force Ceases* [$t = -4.14$, $p < .001$], and *Force Perpendicular to Motion* [$t = -3.17$, $p = .002$], with females scoring significantly lower on those factors. The *t*-test indicated no statistically significant difference for all other factors. This means there were no observed differences in mean composite scores between males and females for those constructs of the FCI.

Table 15
Gender Differences on FCI Factors

FCI Factor	Female	Male	*t*-test	*p*
Vertical and Horizontal Force Combinations	-.066	.071	-1.58	.116
One Direction Force Combinations	.014	-.015	.33	.743
Newton's Third Law	.067	-.072	1.61	.109
Other Situations	-.200	.214	-4.85	.000
Force Ceases	-.170	.182	-4.14	.000
Interpreting Strobe Diagrams	.001	-.001	.03	.980
No Net Force	-.069	.074	-1.66	.097
Force Perpendicular to Motion	-.132	.141	3.17	.002

Testing for Differences on Gender Force Concept Inventory Constructs

No further analysis was conducted for the GFCI, including naming of factors, as small sample size was thought to contribute to difficulties in evaluating the content and common misconceptions associated with the test items of each factor. Once an additional semester of data is collected, a factor analysis will be conducted for the purpose of testing for pedagogy and gender differences of constructs.

CHAPTER 5

INTERPRETATION

THE PEDAGOGY GAP!

SOMEONE OWES WOMEN AN APOLOGY!

For over a decade we have been told that there was no gender difference on the go-to physics assessment, the FCI. Were we listening? Blue and Heller, as well as others, couldn't find the gender gap as far back as 2003. It is alarming that so many headlines have continued to speak of a gender gap, when the only time it can be found is if we ignore the gap in opportunities that seem to have led to the tale-telling pretest gaps. Citizens should be concerned that their girls have gone through their K–12 education under the supervision of adults who have been influenced by these headlines, possibly even patronizing the girls'

> aspirations to study the physical world around them. Why is there no panic over the gender gaps in nursing or education? Headlines collected over time have become part of our inherited generational bias. There is no way to know the depth and breadth of the effects of this "gender gap" fairytale. Like the fairy that trades cash for teeth, the pot of gold, and Santa's crazy all-nighter—it may take years to convince people otherwise. While educators in the US are rethinking how they are contributing to the real gender gaps (fewer females taking high school physics and fewer females taking upper level high school mathematics courses), there is a solution for college physics teachers who are interested in improving all of their students' performance. A new study shows that LEAP pedagogy improves conceptual physics performance—despite the differences that some have tried to finger as a low performance culprit—and explains 30% of the differences in student performance. What a relief it will be for all of the struggling students when they hear that they were not inherently incompetent in physics, after all, and that simply being in a class with a student-centered pedagogy such as LEAP was all they needed for success. What a difference that would make!

Significance of Findings

Pedagogy

Students who were taught physics using LEAP pedagogy significantly outperformed those taught with a traditional approach, even while taking into account pretest performance and the propensity to choose a LEAP course over a traditional

course. This was true when performance was assessed with both the Force Concept Inventory (FCI) and the Gender Force Concept Inventory (GFCI). However, there were four areas of eight FCI constructs in which LEAP students were not doing better than those in the traditional classroom. A low score in any area of the FCI indicated non-Newtonian thinking—thinking that was solidly grounded in a perspective that goes against nature but have aligned with common sense. For *Force Ceases*, *Interpreting Strobe Diagrams*, and *No Net Force*, findings suggest that it was possible that these concepts were comprehended similarly by introductory algebra-based physics students despite being taught with a LEAP or traditional curriculum. It was also possible that pedagogy could not moderate learning of these three areas of the force concept due to the time needed to address the commonsense beliefs that go against nature, reminding us of the "sorry state of affairs" cited by FCI developers (Hestenes et al., 1992). For *Force Perpendicular to Motion*, LEAP students performed significantly less than those in a traditional course. This was an important finding considering that there were multiple activities for directly addressing this concept in the LEAP curriculum.

Gender

Males and females did not differ significantly in physics performance, while taking into account pretest performance and the propensity to choose a LEAP course over a traditional course. This was true when performance was assessed with either the FCI or the GFCI, though on the cusp of significance for the FCI posttest. Students in LEAP, regardless of gender, did better than those in the traditional classroom. Males and females

performed similarly on five of eight FCI constructs. However, there were three areas of eight FCI constructs in which males did significantly better than females. For *Other Situations*, *Force Ceases*, and *Force Perpendicular to Motion*, findings suggested these FCI questions may have been comprehended differently according to gender. This was important for efforts to determine if differences were a manifestation of the assessment item context or differences in experience. These differences could inform the choice of a gender-neutral assessment for evaluation of curricular changes and student progress. Otherwise, use of a biased assessment or interpretation of it at face value leads to misinformed curricular decisions. It is possible that a plethora of findings based on biased assessments comprise the misinformation of the past and led to the common belief that females underperform in the physical sciences in comparison to males.

Pedagogy & Gender

There was a large difference associated with pedagogy, with roughly 30% of the differences attributed to LEAP pedagogy. When asking if the difference in physics performance associated with pedagogy was consistent for males and females, findings indicated that it was consistent. Gender did not moderate pedagogy. In other words, the effect of pedagogy did not change based on someone's gender. Not finding an interaction between pedagogy and gender for posttest analysis or normalized gain analysis, on the FCI or the GFCI, makes the main effect of pedagogy even more generalizable. Not only was it true that LEAP pedagogy was better than traditional on the average, it was true

equally for both genders as confirmed by the lack of interaction. These findings contradict the majority of studies which looked at how interactive engagement methods in physics influenced the gender gap (Madsen et al., 2013).

Recommendations

Pedagogy

The findings of this study suggest that all students, regardless of variations in student background, benefit more from being taught using LEAP curriculum when compared to a traditional approach. This was true despite the metric or test version used here. These findings push back against contradictory works which have found that students gain little despite substantive instruction. The population of introductory physics students would benefit by a strong look at how the LEAP curriculum could be modified for full implementation in a large lecture hall setting. This would provide the greatest opportunity for positive change while minimizing changes to infrastructure.

A next step would be to consider where FCI factors *Force Ceases*, *Interpreting Strobe Diagrams*, and *No Net Force* are addressed in the curriculum and possibly question what changes may address these apparently deep-seated commonsense beliefs. Since traditional students performed statistically better on *Force Perpendicular to Motion*, despite roughly 30% of the overall difference being explained by LEAP

pedagogy, a look at both curricula is particularly important for making improvements to the LEAP curriculum.

Gender

The findings of this study show that males and females do not differ with respect to FCI normalized gain or posttest performance, though gender was near significance for the posttest (p = .052). Males and females did not differ with respect to GFCI normalized gain or posttest performance. All of these findings account for differences in pretest performance as well as the propensity to choose a LEAP course over a traditional course. This questions the large body of PER research stating that a "gender gap" exists and that the FCI is gender-biased. When using LEAP pedagogy to teach introductory algebra-based physics and controlling for the likelihood that student characteristics may have influenced whether or not they would choose a LEAP section, there is no gender difference in overall performance on the FCI or the GFCI. I recommend that the STEM education community look at the K–12 practices that may contribute to differences in pretest performance and unequal access to preparatory experiences, rather than seeing these as deficits inherently attributed to the student and brought to the college classroom. The population of introductory physics students would benefit by a strong look at the specific areas of the FCI in which significant gender differences do exist. A next step is to consider where FCI factors *Other Situations*, *Force Ceases*, and *Force Perpendicular to Motion* are addressed in the curriculum and possibly question the gender related contexts that may contribute to differences in comprehension of the scenarios of the FCI.

This seems particularly important in light of finding that *Other Situations* and *Force Perpendicular to Motion* were significantly different by both gender and pedagogy.

It is also recommended that a GFCI factor analysis be conducted when an additional semester of student data is available so that the factor analysis can be interpretable. This also ensures that sample size is not a barrier to interpreting pedagogy and gender effects. Those results could be used to inform the decision of whether or not the GFCI should be used, in lieu of the FCI, to measure introductory algebra-based physics students' performance.

Synthesis of the Findings

About 35% of the variance in FCI performance was accounted for by the model in which pedagogy and gender were the independent variables, controlling for predicted group membership and pretest performance. Males and females did not differ significantly in physics performance. These findings aligned with Blue and Heller (2003), who found no difference in male and female FCI performance when compared on high school GPA, class level, and other relevant background characteristics. Findings suggested that performance on some FCI constructs were different according to gender, which was the foundation for Laura McCullough's work in redesigning the FCI contexts (McCullough & Foster, 2001). Questions specific to *Other Situations* and *Force Perpendicular to Motion* were significantly different by both gender and pedagogy. LEAP pedagogy is better for both males and females. Because the large difference in

performance attributed to pedagogy does not change based on a student's gender, the effect of pedagogy is even stronger and more generalizable.

Limitations

Engineering Technologies (previously termed Industrial Technology) students at TTU might have been enrolled in traditional sections of PHYS2010 more than LEAP sections because of scheduling conflicts due to blocks of courses in the major. When students were asked by departmental survey why they switched between sections of PHYS2010, schedule timing was found to drive student selection of traditional or LEAP sections.

Two instructors for LEAP sections of PHYS2010 did not teach traditional sections of PHYS2010. It could have been argued that differences found in students were partly attributed to the instructor rather than the pedagogy used by the instructor. Though there was the potential for concern due to multiple instructors being used in analysis, the use of many different instructors meant that the analysis was not confounded, however uncontrolled.

Due to the small size of the GFCI sample, results were preliminary and indicated the need for additional cases. Sample size was also reduced in this study by the lack of identifying information necessary to match pretests with the corresponding posttests in some entire classes. This made course section comparisons within a semester impossible for some semesters.

Not only was it possible that selected propensity score covariates were inappropriate, but it was reasonable to expect that some relevant covariates were not included. For this study, some covariates were not included in propensity scoring because they were not available. For instance, parental STEM career was desired as a dichotomous student covariate. Though that information existed, it was not feasible to collect it from individuals. For reasons such as this, parental STEM career and other variables were not used as covariates in this study.

Discussion

This study is situated in the critical discourse of Chambers (2009), Gutiérrez (2008), Ladson-Billings (2006), and Lather (2012) in that achievement gap focus is a deficit model that further promotes stereotypes, placing the burden of equity on the marginalized. The enduring stereotype of low performance of females in physics, without a strong objective look at the instructional methods and assessment choices that have helped create that perceived difference, has arguably continued as a result of a lack of headlines touting a contrary finding. The way STEM learning is measured influences our interpretation of student performance. In short, this study found no gender gap in physics performance. Rather than gender accounting for differences in student performance, LEAP pedagogy explains nearly a third of those differences. Furthermore, the large differences associated with pedagogy do not hinge upon being male or female. The gender differences were even more insignificant when measured using the GFCI.

STEM pedagogies help create the culture of STEM learning (Shulman, 2005), and LEAP pedagogy was better for both males and females. As Zohar and Sela (2003) pointed out, the success of girls in physics is a remarkable accomplishment considering that traditional teaching methods often are not equitable. Like the pedagogical tenants found to increase FCI gains in Hake's landmark study (1998), the contexts of learning and the resulting culture in a LEAP classroom supports academic peer talk and interaction that is not characteristic of the traditional classroom. Students are responsible for sharing disagreement with peers, where all conceptions are valued. This LEAP culture of learning may mitigate the gender threat of being in a setting that is perceived to be better suited for males, so that females feel safe in verbalizing their conceptions and thus are provided the opportunity for changes in thinking. The fact that LEAP pedagogy was better for males as well as females further supports the idea that the culture of the classroom and the choice of contexts are important for accurately evaluating all students' performance.

Teacher education programs have long supported the tenant of special education in which there is a focus on similarities among learners rather than their differences (Lewis & Doorlag, 2011). For STEM education, there seems to be a focus on gender differences rather than commonalities. This dichotomy in teaching philosophy merits critical consideration. An important implication for this work was addressing the enduring stereotype of female underperformance in physics by exposing the instructional methods that have helped create a perceived difference. Based on the findings of this work, the conversation about gender gaps should stop before more headlines get in the minds of our students and the pre-service teachers who will teach them. The focus should

turn to pedagogy gaps. This analysis offers a taste of strong objectivity which shows LEAP pedagogy and curriculum to be highly effective in comparison to other pedagogical choices, particularly considering that teaching methods research rarely yields these large main effects while using a similar statistical treatment. Using that knowledge to effect a deliberate restructure of the culture of academia, including physics classroom practices, is necessary to "break down the barriers constraining women's participation and effectiveness" (Bilimoria et al., 2008) and influencing the women in science to stay in science. Those who have investigated gender bias of physics assessments have started the conversation (Dietz et al., 2012). Laura McCullough also advanced that conversation by creating a female-centric version of one such assessment. Addressing bias in assessments and the need for gender-neutral pedagogies should be considered a priority in institutional equity movements.

Being powerblind to the differences in opportunities and practices in the K–12 setting only contributes to deficit thinking (Chambers, 2009; Gutiérrez, 2008; Ladson-Billings, 2006; & Lather, 2012). Attributing variance in performance to gaps in ability, preparation, and background without adjusting for such differences would contribute to the idea that being underprepared is a choice. By reporting findings that have not been adjusted to account for the variables that intuitively relate to group assignment or the performance measure, there is an objectivity deficit—possibly situated at the shallow end of the empirical pool. I am not suggesting that all studies which investigate science performance as a function of mathematics ability, for example, lack the strong objectivity that Sandra Harding (1993a) called for in empirical research. I am instead suggesting that

such studies are profoundly important in adding to the discourse on STEM education and should be as unassuming as a medical trial.

To illustrate the heavy implications of using a deficit ideology to explain the factors contributing to gender differences in physics, I will draw on a medical analogy. It would be faulty and misguiding to report differences in overall health of children without addressing poor water quality, lack of access to adequate nutrition, or limited access to healthcare. And at no point would the findings of such a study be presented in a way that suggested that some children are of poor health because they choose to eat an unbalanced diet or choose to live in a community that has little access to pediatric care. In the case of physics education research, there is a reasonable expectation that mathematics ability influences understanding of physics. A deficit model of thinking would explain low ability in physics as an inherent lack of aptitude or failure to seek opportunities to improve ability in mathematics (Chambers, 2009; Gutiérrez, 2008; Ladson-Billings, 2006; & Lather, 2012), while Hyde et al. (2008) suggested that such deficits can be attributed to course goals rather than the student. Rather than being powerblind to hegemonic schooling practices, I am suggesting that educators, administrators, and advisors create equal opportunities and adequate college-preparatory STEM coursework and experiences for every student.

REFERENCES

Adamson, S., Banks, D., Burtch, M., Cox, F., Judson, E., Turley, J., Benford, R., & Lawson, A. (2003). Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching, 40*(10), 939–957.

Asher, N. (2002). (En)gendering a hybrid consciousness. *Journal of Curriculum Theorizing, 18*(4), 81–92.

Austin, P. (2011a). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Austin, P. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*(2), 150–161.

Baram-Tsabari, A., & Yarden, A. (2008). Girls' biology, boys' physics: Evidence from free-choice science learning settings. *Research in Science and Technological Education, 26*(1), 75–92.

Baram-Tsabari, A., & Yarden, A. (2009). Identifying meta-clusters of students' interest in in science and their change with age. *Journal of Research in Science Teaching, 46*(9), 999–1022.

Baram-Tsabari, A., & Yarden, A. (2011). Quantifying the gender gap in science interests. *International Journal of Science and Mathematics Education, 9*(3), 523–550.

Baron, R., & Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology, 51*(6), 1173–1182.

Baron-Cohen, S. (2007). Sex differences in mind: Keeping science distinct from social policy. In S. Ceci, & W. Williams, (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 159–172). Washington, DC: American Psychological Association.

Barton, A., Tan, E., & Rivet, A. (2008). Creating hybrid spaces for engaging school science among urban middle school girls. *American Educational Research Journal, 45*(1), 68–103.

Bazzul, J. (2013). Emancipating subjects in science education: taking a lesson from Patti Lather and Jacques Rancière. *Cultural Studies of Science Education, 8*(1), 245–251.

Beede, D., Julian, T., Langdon, D., McKittrick, G., Khan, B., & Doms, M. (2011). Women in STEM: A gender gap to innovation (ESA Issue Brief# 04-11). *Washington, DC: US Department of Commerce*.

Bensimon, E. (2005). Closing the achievement gap in higher education: An organizational learning perspective. *New Directions for Higher Education, 131*, 99–111

Bilimoria, D., Joy, S., & Liang, X. (2008). Breaking barriers and creating inclusiveness: Lessons of organizational transformation to advance women faculty in academic science and engineering. *Human Resource Management*, *47*(3), 423–441.

Bomer, R., Dworin, J., May, L., & Semingson, P. (2008). Miseducating teachers about the poor: A critical analysis of Ruby Payne's claims about poverty. *The Teachers College Record*, *110*(12), 2497–2531.

Bower, B. (1998). Objective visions: Historians track the rise and times of scientific objectivity. *Science News*, *154*(23), 360–362.

Blue, J., & Heller, P. (2003). Using matched samples to look for sex differences. *Proceedings of Physics Education Research Conference 2003*, Madison, WI. *720*(1), 45–48.

Britzman, D. (1994). Is there a problem with knowing thyself? Toward a poststructuralist view of teacher identity. In Timothy Shanahan (Ed.), *Teachers thinking, teachers knowing: Reflections on literacy and language education* (pp. 53–75). Urbana, IL: National Council of Teachers of English.

Buse, K., Bilimoria, D., & Perelli, S. (2013). Why they stay: women persisting in US engineering careers. *Career Development International*, *18*(2), 139–154.

Carlson, J., & Kwon, H. (2006). *Delivering integrative science technology engineering and mathematics [STEM] education: Technology teachers' experiences in a summer camp.* Unpublished manuscript, Departments of Technology Education and Integrative STEM Education, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Carrell, S., Page, M., & West, J. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, *125*(3), 1101–1144.

Chambers, T. (2009). The "receivement gap": School tracking policies and the fallacy of the "achievement gap". *The Journal of Negro Education*, *78*(4), 417–431.

Clouston, S. (2013). Propensity score matching and longitudinal research designs: Counterfactual analysis using longitudinal data. In D. Kuh, R. Cooper, R. Hardy, M. Richards, & Y. Ben-Shlomo (Eds.), *A Life Course Approach to Healthy Ageing* (pp. 109–117). New York, NY: Oxford University Press.

Coletta, V. (2015). *Thinking in Physics*. Hoboken, NJ: Pearson Education.

Coletta, V., Phillips, J., & Steinert, J. (2012). FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects. In *AIP Conference Proceedings* (Vol. 1413, p. 23).

Commeyras, M., & Alvermann, D. (1996). Reading about women in world history textbooks from one feminist perspective. *Gender and Education, 8*(1), 31–48.

Confrey, J., & Lachance, A. (2000). Transforming teaching experiments through conjecture-driven research design. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 231–265). Mahwah, NJ: Lawrence Erlbaum Associates.

Cronin Jones, L. (2003). Are lectures a thing of the past? Tips and techniques for success. *Journal of College Science Teaching, 32*(7), 453–457.

Crotty, M. (2003). *The foundations of social research. Meaning and perspective in the research process.* London: Sage.

Dancy, M. (2004). The myth of gender neutrality. In *AIP Conference Proceedings* (Vol. 720, p. 31).

Deleuze, G., & Guattari, F. (1987). *A thousand plateaus: Capitalism and schizophrenia*. London: Burns & Oates.

Dietz, R., Pearson, R., Semak, M., & Willis, C. (2012). Gender bias in the force concept inventory?. In *AIP Conference Proceedings* (Vol. 1413, p. 171).

Dudley-Marling, C. (2007). Return of the deficit. *Journal of Educational Controversy, 2*(1).

Eisenhart, M., & Holland, D. (1990). *Educated in romance: Women, achievement, and college culture*. Chicago: University of Chicago Press.

Else-Quest, N., Hyde, J., & Linn, M. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127.

Engelhardt, P. (2009). An introduction to classical test theory as applied to conceptual multiple-choice tests. *Getting Started in PER, 2*(1).

Eris, O., Chachra, D., Chen, H., Sheppard, S., Ludlow, L., Rosca, C., Bailey, T., Toye, G. (2010). Outcomes of a longitudinal administration of the persistence in engineering survey. *Journal of Engineering Education, 99*(4), 371–395.

Equal Opportunity Commission of Hong Kong. (2001). *Stereotypes in textbooks and teaching materials in Hong Kong*: *A literature review.* Human Rights Education in Asian Schools, *6.*

Farnham-Diggory, S. (1994). Paradigms of knowledge and instruction. *Review of Educational Research, 64*(3), 463–477.

Goldberg, F., Otero, V., & Robinson, S. (2010). Design principles for effective physics instruction: A case from physics and everyday thinking. *American Journal of Physics, 78*(12), 1265–1277.

Gorski, P. (2008). The myth of the culture of poverty. *Educational Leadership, 65*(7), 32–36.

Gorski, P. (2010). The evolution of a pro-feminist. In M. Adams, W. Blumenfeld, C. Castañeda, H. Hackman, M. Peters, & X. Zúñiga (Eds.), *Readings for diversity and social justice* (2nd ed, pp. 356–357). New York: Taylor & Francis.

Gorski, P. (2011). Unlearning deficit ideology and the scornful gaze: Thoughts on authenticating the class discourse in education. In R. Ahlquist, P. Gorski, & T. Montaño (Eds.), *Assault on kids: How hyper-accountability, corporatization, deficit ideology, and Ruby Payne are destroying our schools* (pp. 152–175). New York, NY: Peter Lang.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320*, 1164–1165.

Gutiérrez, R. (2008). A "gap-gazing" fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education, 39*(4), 357–364.

Hake, 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*(1), 64–74.

Hake, R. (2006). Possible palliatives for the paralyzing pre/post paranoia that plagues some PEP's. *Journal of Multidisciplinary Evaluation, 3*(6), 60–71.

Hake, R. (2007). Design-Based Research in Physics Education: A Review. In Kelly, A., Lesh, R., & Baek, J. (Eds.), *Handbook of design research methods in mathematics, science, and technology education*, (pp. 493–508).

Halloun, I, & Hestenes, D. (1985a). The initial knowledge state of college physics students. *American Journal of Physics, 53*(11), 1043–1055.

Halloun, I, & Hestenes, D. (1985b). Common sense concepts about motion. *American Journal of Physics, 53*(11), 1056–1065.

Halloun, I., & Hestenes, D. (1996). *The search for conceptual coherence in FCI data*. Unpublished manuscript, Department of Physics and Astronomy, University of Arizona.

Harding, S. (1993a). Rethinking standpoint epistemology: What is "strong objectivity"? In L. Alcoff & E. Potter (Eds.), *Feminist Epistemologies* (pp. 49–82). New York: Routledge.

Harding, S. (1993b). Introduction: Eurocentric scientific illiteracy—A challenge for the world community. In S. Harding (Ed.), *The "racial" economy of science: Toward a democratic future* (pp. 1–22). Bloomington: Indiana University Press.

Harding, S. (1998). *Is Science Multicultural?* Bloomington: Indiana University Press.

Heller, P., & Huffman, D. (1995). Interpreting the Force Concept Inventory: A reply to Hestenes & Halloun. *The Physics Teacher, 33*(8), 503, 507–511.

Hestenes, D., & Halloun, I. (1995). Interpreting the Force Concept Inventory: A response to Huffman & Heller. *The Physics Teacher, 33*(8), 502, 504–506.

Hestenes, D., & Halloun, I. (1996). *The search for conceptual coherence in FCI data*. Unpublished manuscript, Department of Physics and Astronomy, University of Arizona.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141–151.

Hewlett, S., Luce, C., & Servon, L. (June 2008). Stopping the exodus of women in science. *Harvard Business Review*, 22–24.

Hines, M. (2007). Do sex differences in cognition cause the shortage of women in science? In S. Ceci, & W. Williams, (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 101–112). Washington, DC: American Psychological Association.

Hoffman, L. (2002). Promoting girls' interest and achievement in physics classes for beginners. *Learning and Instruction, 12*(4), 447–465.

Hubbard, R. (2003). Who's Helen Keller? Do children's books distort the truth of Helen Keller's life? *Teaching Tolerance, 24*, 26–32.

Hyde, J., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139–155.

Hyde, J., Lindberg, S., Linn, M., Ellis, A., Williams, C. (2008). Gender similarities characterize math performance. *Science, 321*(5888), 494–495.

International Technology Education Association (2000). *Standards for technological literacy: Content for the study of technology*. Reston, VA: Author. Retrieved from http://www.iteea.org/TAA/PDFs/xstnd.pdf.

Kachigan, S. (1991). *Multivariate Statistical Analysis: A Conceptual Introduction.* New York: Radius Press.

Kelly, A. (1978). *Girls and science: An international study of sex differences in school science achievement*. Stockholm: Almqvist & Wiksell International.

Kimura, D. (2007). "Underrepresentation" or misrepresentation? In S. Ceci, & W. Williams, (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 39–46). Washington, DC: American Psychological Association.

Kost, L., Pollock, S., & Finkelstein, N. (2009). Characterizing the gender gap in introductory physics. *Physical Review Special Topics-Physics Education Research, 5*(1), 010101.

Kost-Smith, L. (2011). Characterizing, modeling, and addressing gender disparities in introductory college physics (Doctoral dissertation). Retrieved from *Dissertation Abstracts International, 72-07* (AAT 3453741).

Krapp, A. (2000). Interest and human development during adolescence: An educational-psychological approach. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 109–128). London: Elsevier.

Kurzman, C., Ghoshal, R., Gibson, K., Key, C., Roos, M. & Wells, A. (2014). Powerblindness. *Sociology Compass, 8*(6), 718–730.

Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education, 11*(1), 7–24.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher, 35*(7), 3–12.

Lasry, N., Rosenfield, S., Dedic, H., Dahan, A., & Reshef, O. (2011). The puzzling reliability of the Force Concept Inventory. *American Journal of Physics, 79*(9), 909–912.

Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange, 17*(4), 63–84.

Lather, P. (1991). *Getting smart: Feminist research and pedagogy with/in the postmodern.* New York: Routledge.

Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *The Sociological Quarterly, 34*(4), 673–693.

Lather, P. (2004a). Critical inquiry in qualitative research: Feminist and poststructural perspectives: Science "after truth". In K. deMarrais & S. Lapan (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 203–216). Mahwah, NJ: Lawrence Eribaum Associates.

Lather, P. (2004b). Scientific research in education: A critical perspective. *British Educational Research Journal, 30*(6), 759–772.

Lather, P. (2005). Scientism and scientificity in the rage for accountability: A feminist deconstruction. Paper presented at the First International Congress of Qualitative Inquiry, Urbana-Champaign, Illinois.

Lather, P. (2012). The ruins of neo-liberalism and the construction of a new (scientific) subjectivity. *Cultural Studies of Science Education, 7*(4), 1021–1025.

Lewis, R., & Doorlag, D. (2011). *Teaching Students with Special Needs in General Education Classrooms (Custom ed).* Upper Saddle River, NJ: Prentice Hall.

Lorenzo, M., Crouch, C., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, *74*(2), 118–122.

Lyotard, J. (1984). *The postmodern condition: A report on knowledge* (Vol. 10). University of Minnesota Press.

MacIsaac, D., & Falconer, K. (2002). Reforming physics instruction via RTOP. *The Physics Teacher*, *40*(8), 479–485.

Madsen, A., McKagan, S., & Sayre, E. (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*(2), 020121.

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

McClaren, P. (2003). *Life in schools: An introduction to critical pedagogy in the foundations of education* (4th ed.). Boston: Allyn & Bacon.

McCullough, L., & Foster, T. (2001). A gender context for the Force Concept Inventory. In *meeting of the American Association of Physics Teachers, San Diego, CA*.

McCullough, L. (2002). Gender, Math, and the FCI. Paper presented at Physics Education Research Conference 2002, Boise, Idaho. Retrieved November 30, 2013, from http://www.compadre.org/Repository/document/ServeFile.cfm?ID=4328&DocID=1150.

McCullough, L. (2004). Gender, context, and physics assessment. *Journal of International Women's Studies*, *5*(4), 20–30.

McCullough, L. (2011). Gender differences in student responses to physics conceptual questions based on question context. ASQ STEM Agenda Conference Proceedings. University of Wisconsin-Stout, Menomonie, WI.

McDermott, L., & Redish, E. (1999). Resource letter: PER-1: Physics education research. *American journal of physics*, *67*(9), 755–767.

Mertler, C., & Vannatta A. (2005). *Advanced and Multivariate Statistical Methods* (4th ed.). Glendale, CA: Pyrczak Publishing.

Morrell, P., Flick, L., & Wainwright, C. (2004). Reform teaching strategies used by student teachers. *School Science and Mathematics*, *104*(5), 199–213.

Mulford, D., & Robinson, W. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education*, *79*(6), 739–744.

National Research Council. (2002). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: The National Academies Press.

National Research Council. (2006). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future.* Committee on Science, Engineering and Public Policy. Washington, D.C.: National Academy Press.

National Research Council (2012). *A Framework for K-12 Science Education: Practices, crosscutting concepts, and core ideas.* Washington, DC: National Academies Press.

Noack, A., Antimirova, T., & Milner-Bolotin, M. (2009). Student diversity and the persistence of gender effects on conceptual physics learning. *Canadian Journal of Physics*, *87*(12), 1269–1274.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.

Pannabecker, J. (2002). Integrating technology, science, mathematics at Napolean's school for industry, 1806–1815. *Journal of Technology Education*, *14*(1), 51–64.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). New York: Cambridge University Press.

Pibern, M., & Sawada, D. (2000). *Reformed Teaching Observation Protocol (RTOP) Reference Manual* (ACEPT Technical Report No. IN00-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Pillow, W. (2000). Deciphering attempts to decipher postmodern educational research. *Educational Researcher, 29*(5), 21–24.

Pollock, S., Finkelstein, N., & Kost, L. (2007). Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physical Review Special Topics-Physics Education Research*, *3*(1), 010107.

Razali, N., & Wah, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Riegle-Crumb, C., & Moore, C. (2014). The gender gap in high school physics: considering the context of local communities. *Social Science Quarterly*, *95*(1), 253–268.

Robenstine, C. (1992). French colonial policy and the education of women and minorities: Louisiana in the early eighteenth century. *History of Education Quarterly, 32*(2), 193–211.

Rosenbaum, P., & Rubin, D. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rosenbaum, P., & Rubin, D. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, 45*(2), 212–218.

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

Rudner, L., & Peyton, J. (2006). Consider propensity scores to compare treatments. *Practical Assessment, Research & Evaluation, 11*(9).

Rutherford, F., & Ahlgren, A. (1989). *Science for all Americans*. New York: Oxford University Press. Retrieved from http://www.project2061.org/publications/sfaa/online/chapter3.htm.

Sabella, M., & Van Duzor, A. (2013). Cultural toolkits in the urban physics learning community. In *American Institute of Physics Conference Series* (Vol. 1513, pp. 34–37).

Sanders, M. (2006). A rationale for new approaches to STEM education and the STEM education graduate programs. Paper presented at the 93rd Mississippi Valley Technology Teacher Education Conference. Nashville, TN.

Sanders, M. (2009). STEM, STEM education, STEMmania. *The Technology Teacher, 68*(4), 20–26.

Sawada, D., Piburn, M., Judson, E., Turley, J., Falconer, K., Benford, R. & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics, 102*(6), 245–253.

Schwalbe, M., Godwin, S., Holden, D., Schrock, D., Thompson, S. & Wolkomir, M. (2000). Generic processes in the reproduction of inequality: An interactionist analysis. *Social Forces, 79*(2), 419–452.

Schwartz, D., & Bransford, J. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522.

Servon, L., & Visser, M. (2011). Progress hindered: The retention and advancement of women in science, engineering and technology careers. *Human Resource Management Journal, 21*(3), 272–284.

Shore, Z. (2000). Girls learning, women teaching: Dancing to different drummers. *Educational Studies 31*(2), 132–145.

Shulman, L. (2005). Signature pedagogies in the professions. *Daedalus, 134*(3), 52–59.

Sipe, L., & Constable, S. (1996). A chart of four contemporary research paradigms: Metaphors for the modes of inquiry. *Taboo: The Journal of Culture and Education, 1,* 153–163.

Spelke, E., & Grace, A. (2007). Sex, math, and science. In S. Ceci, & W. Williams, (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 57–67). Washington, DC: American Psychological Association.

Spring, J. (2004). *Deculturalization and the struggle for equality: A brief history of the education of dominated cultures in the United States* (4th ed.). Boston: McGraw-Hill Higher Education.

St. Pierre, E. (2000). The call for intelligibility in postmodern educational research. *Educational Researcher, 29*(5), 25–28.

St. Pierre, E. (2002). Comment: "Science" rejects postmodernism. *Educational Researcher, 31*(8), 25–27.

Thornton, R., Kuhl, D., Cummings, K., Marx. J. (2009). Comparing the Force and Motion Conceptual Evaluation and the Force Concept Inventory. *Physical Review Special Topic-Physics Education Research, 5*(1), 1–8.

Tozer, S., Violas, P., & Senese, G. (Eds.). (2002). *School and society: Historical and contemporary perspectives* (4th ed.). Boston: McGraw-Hill Higher Education.

Tyack, D. (1974). *The one best system: A history of American urban education.* Cambridge, MA: Harvard University Press.

Wainwright, C., Flick, L., & Morrell, P. (2003). The development of instruments for assessment of instructional practices in standards-based teaching. *The Journal of Mathematics and Science: Collaborative Explorations, 6,* 21–46.

Wainwright, C., Flick, L., & Morrell, P., Schepige, A. (2003). Observation of reform teaching in undergraduate level mathematics and science courses. *School Science and Mathematics, 104*(7), 322–335.

Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics, 78*(10), 1064–1070.

Witte, R., & Witte, J. (2007). *Statistics* (8th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Wuensch, K. (2014). *Binary Logistic Regression with SPSS.* Retrieved from http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.PDF

Zohar, A., & Bronshtein, B. (2005). Physics teachers' knowledge and beliefs regarding girls' low participation rates in advanced physics classes. *International Journal of Science Education, 27*(1), 61–77.

Zohar, A., & Sela, D. (2003). Her physics, his physics: Gender issues in Israeli advanced placement physics classes. *International Journal of Science Education*, *25*(2), 245–268.

BIBLIOGRAPHY

Bao, L., & Redish, E. (2001). Concentration analysis: A quantitative assessment of student states. *American Journal of Physics, 69,* S45–S53.

Cohen, P., Cohen, J., Aiken, L., & West, S. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34*(3), 315–346.

Cronbach, L., & Furby, L. (1970). How we should measure "change"—or should we? *Psychological Bulletin, 74,* 68–80.

Dehejia, R., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association, 94*(448), 1053–1062.

Glass, G., & Hopkins, K. (1996). *Statistical Methods in Education and Psychology* (3rd ed.). Boston: Allyn & Bacon.

Gu, X., & Rosenbaum, P. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics, 2*(4), 405–420.

Hill, J., & Reiter, J. (2006). Interval estimation for treatment effects using propensity score matching. *Statistics in Medicine, 25*(13), 2230–2256.

Masicampo, E., & Lalande, D. (2012). A peculiar prevalence of *p* values just below. 05. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2271–2279.

Ming, K., & Rosenbaum, P. (2000). Substantial gains in bias reduction from matching with a variable number of controls. *Biometrics, 56*(1), 118–124.

Newspaper Headlines Template. (n.d.). Retrieved April 19, 2015, from http://www.presentationmagazine.com/newspaper-headlines-template-9437.htm

Nieminen, P., Savinainen, A., & Viiri, J. (2012). Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning. *Physical Review Special Topics-Physics Education Research, 8*(1), 010123.

Normand, S., Landrum, M., Guadagnoli, E., Ayanian, J., Ryan, T., Cleary, P., & McNeil, B. (2001). Validating recommendations for coronary angiography following an acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology, 54*(4), 387–398.

Parsons, E., Miles, R., & Peterson, M. (2011). High school students' implicit theories of what facilitates science learning. *Research in Science & Technological Education, 29*(3), 257–274.

Rosenbaum, P. (2002). *Observational studies* (2nd ed.). New York, NY: Springer-Verlag.

Rosenbaum, P., & Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33–38.

Semak, M., & Dietz, R. (2014). Aspects of Factor Analysis Applied to the Force Concept Inventory. Paper presented at Physics Education Research Conference 2014, Minneapolis, MN.

Semak, M., Pearson, R., Dietz, R. & Willis, C. (2010). Factor Analysis and Question Categorization in the Force Concept Inventory. In *meeting of the American Association of Physics Teachers*, Washington, D.C..

St. Pierre, E. (2011). Post-qualitative research: The critique and the coming after. In N. Denzin & Y. Lincoln (Eds.), *The handbook of qualitative research* (4th ed., pp. 611–635). Thousand Oaks, CA: Sage.

Willis, C., Semak, M., & Dietz, R. (2009). Factor Analysis and the Force Concept Inventory. In *meeting of the American Association of Physics Teachers,* Ann Arbor, MI.

APPENDIX

CODING OF MAJORS & PROGRAMS

Category for Coding	Program
Engineering	Engineering Technology
	Mechanical Engineering
Pre-Professional	Agriculture
	Biology
	Chemistry
	General Health Studies
	Nursing-Lower Division
	Physics
	Pre-Dentistry
	Pre-Medicine
	Pre-Optometry
	Pre-Pharmacy
Exercise Sciences	Exercise Sci, PhysEd, Wellness
	Pre-Occupational Therapy
	Pre-Physical Therapy
Other Majors	Accounting
	Basic Business
	Communication
	Computer Science
	English
	Environmntl & Sustain Studies
	Foreign Languages
	General Curriculum
	Geosciences
	Human Ecology
	Interdisciplinary Studies
	Multidisciplinary Studies
	Music
	Pre-Dental Hygiene

VITA

Twanelle Deann Walker Majors was born in McMinnville, Tennessee, on July, 21, 1973. She attended Westwood Kindergarten and the elementary and middle schools of the Warren County School District. She graduated from Warren County High School in May 1991. The following August she entered University of Tennessee Chattanooga and in May 1996 received the degree of Bachelor of Science in Pre-Medicine: Biology. Within that time she gave birth to two daughters, Treeah Baili and Chloe Très. She entered Tennessee Technological University in August 1998 and received Chemistry and Biology 7–12 endorsements as well as a Master of Arts degree in Secondary Science Curriculum and Instruction in May 2005. During that time she gave birth to a son and a daughter, Rily Tru and Ginger Teer Lu. She entered Tennessee Technological University in August 2010 and received a Specialist in Education degree in Curriculum and Instruction in May 2011. The following August she entered Tennessee Technological University and received a Doctor of Philosophy degree in Exceptional Learning with a concentration in Science, Technology, Engineering, and Mathematics (STEM) in May 2015.