A COMPARISON OF LEARNING IN DUAL CREDIT

INTRODUCTORY COLLEGE PHYSICS

by

Eric L. Thies

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in STEM Education

Boise State University

December 2014

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Eric L. Thies

Thesis Title: A Comparison of Learning in Dual Credit Introductory College Physics

Date of Final Oral Examination: 05 August 2014

The following individuals read and discussed the thesis submitted by student Eric L. Thies, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Keith W. Thiede, Ph.D.	Chair, Supervisory Committee
Laurie Cavey, Ph.D.	Member, Supervisory Committee
Dmitri A. Tenne, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Keith W. Thiede, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

ACKNOWLEDGMENTS

To my committee members. Dr. Tenne, thank you for giving me the time to work on my thesis, even when I should have been shooting big lasers at small materials. Dr. Cavey, thank you for watching my back, even when I didn't know it needed watching. Dr. Thiede, thank you for stepping up at the last minute to help me finish.

To the graduate college office for giving me encouragement at just the right time.

To my wife. Thank you for suffering through my rants and raves, my late nights, early mornings and long days. But mostly for giving up your summer so that I could get this done. Next summer, we do what you want.

ABSTRACT

Dual credit courses have been offered for over fifty years and have helped students save time and money during their college education. However, little has been done to study the quality of the dual credit courses themselves. The literature is unclear about whether students in dual credit programs learn the same material as the students enrolled in the same course at the university level.

The purpose of my study was to determine whether students in a concurrent enrollment introductory physics course achieve the same knowledge growth as university students enrolled in the same physics course. I used the Force Concept Inventory (FCI) as a measure of students' knowledge. The FCI was given as both a pre-instruction and post-instruction assessment to both the high school and university students and I used a 2 x 2 analysis of variance to compare the two groups at the two different times.

I found that both the high school group and the university group showed significant growth from pre- to post-instruction. I also found that the high school group scored significantly higher than the university group on both the pre-instruction and post-instruction FCI and the high school students showed marginally greater growth. Any conclusions drawn from my study should be tempered with the understanding that the FCI only addresses a portion of the curriculum covered in each course, the sample size was small, including only one high school and one university class, and there was no

consideration for long-term retention of knowledge. However, my conclusion is that dual credit courses may offer students the same knowledge as regular university courses.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

I have noticed increased social and political pressure to increase the number of dual enrollment classes offered to students at the high school level. Some states have recently passed laws funding dual credit courses (Idaho State Department of Education, n.d.). National magazines are ranking schools based on how many students are participating in dual credit courses (Moris, 2014). With the increased pressure to increase dual credit offerings, it is important to determine if dual credit courses are benefiting the students who take them.

I have adopted Oregon Department of Education's definition of "dual credit," which defines a dual credit program as any program that "award(s) secondary and postsecondary credit for a course offered in a high school during regular school hours" (Oregon Department of Education, 2014). In my review of literature, I found several different dual credit programs with three basic structures. The first structure for dual credit has students attend a high school course that is hosted at a college or university where a college or university professor teaches the course and assigns the grade for both the high school and college courses (Edmunds et al., 2010). The second structure for dual credit has high school teachers, who have been vetted by a university or college, teach students in a high school classroom, and the high school teacher is responsible for assigning the students' high school and university grades (Juarez-Coca, 2012). The third structure has high school students taking a test, the scores on which are used by universities to determine college credit (Juarez-Coca, 2012).

I focused my attention on two types of programs that I thought could work in a typical high school environment: Advanced Placement (AP) programs from the College Board, and Concurrent Enrollment (CE) courses offered by a local university. As I reviewed the literature, I noticed that very little had been done to study the students' acquisition of knowledge in dual credit programs. Many comparisons had been drawn between the different types of dual credit programs (Bryant, 2001; Byrd, Elligton, Gross, Jago, & Stern, 2007; Emunds et al., 2010; Farkas & Duffett, 2009; Juarez-Coca, 2012; Marshall, & Andrews, 2002; Tinberg & Nadeau, 2011), but none of these studies focused on measuring students' acquisition of knowledge.

Statement of the Problem

The question I had was about the quality of education received by the high school students measured by the acquisition of knowledge. The literature about dual credit courses reveals very little about the knowledge acquired during a dual credit course, and focuses instead on dual credit programs providing a means for students to reduce the amount of time and money spent in a post-secondary program (Bryant, 2001; College Board AP, 2009; Hébert, 2001; Juarez-Coca, 2012).

Students who participate in dual credit courses are more likely to enroll in college and graduate in less time than students who do not. Dual credit programs reduce a student's time in college by as much as two years, resulting in a significant reduction in tuition paid (Marshal & Andrews, 2002). I argue that the time and money that can be saved are only two considerations to look at when examining dual credit programs.

Perhaps the most important consideration is whether dual credit courses produce the same learning opportunities for students.

There is evidence to suggest that high school instructors cannot produce the same results as post-secondary instructors and that the students receive insufficient opportunity to learn at the high school (Bryant, 2001; Hébert, 2001). Thus, taking dual credit courses may be saving time and money at the expense of learning. I think this is an important thing to investigate, so I pose the question, do high school students enrolled in dual credit courses acquire the same knowledge as the post-secondary students enrolled in the same college or university course?

Purpose of the Study

A survey of high school graduates revealed that some high school juniors and seniors felt their final years in high school was wasted time (Tinberg & Nadeau, 2011). Tinberg and Nadeau (2011) came to the conclusion that the "wasted time" was a result of two things. First, the high school seniors had met all the academic requirements for graduation; and second, they found no challenge in the additional courses they had to attend to receive their high school diploma. To provide high school seniors with a challenge that would engage and motivate the students, educators created dual credit courses (Marshall & Andrews, 2002; Tinberg & Nadeau, 2011; Watt-Malcolm, 2011).

In Concurrent Enrollment (CE) courses, teachers are given the responsibility to determine the students' grades for the course at both the high school and college or university (Juarez-Coca, 2012). I argue that the reliance on the high school teachers to assess student learning puts more emphasis on the ability of the CE teacher to create an academically challenging course. Without common standards or assessments of students, there is no evidence that students are learning comparable material at the high school and university.

Students who participate in dual credit courses are receiving college credits and therefore reducing the amount of time and money they spend on their college degree (Hébert 2001; Juarez-Coca, 2012; College Board AP, 2009). However, little is known about the comparability of learning experience in dual credit courses and typical university courses. With the importance put on dual credit courses and the expected savings of time and money, it seems important to look at the ability of dual credit programs to provide students the same opportunity to gain knowledge as a traditional university course.

This study sought to answer the following question: Can high school students taking CE courses acquire the same knowledge as the students enrolled in the same course at the college or university?

Significance of Study

The majority of research pertaining to the knowledge gained during dual credit courses are studies that addressed issues of long-term success in future course work. Students who take introductory courses through a dual enrollment program tend to have a higher rate of graduation in post-secondary programs, but have lower grades in upper division courses (North & Jacobs, 2010). With the mixed results provided by North and Jacobs (2010), I wanted to investigate whether CE courses provide students with a knowledge base that is comparable to the students' who take the course at the college or university. In particular, I designed my study to examine whether CE students enrolled in an introductory physics course could acquire the same knowledge or conceptual development of forces as university students enrolled in the same course.

CHAPTER II: REVIEW OF LITERATURE

Dual credit courses are accepted by many as beneficial to students (Bryant, 2001; College Board AP, 2009; Emunds et al., 2010; Hébert, L., 2001; International Baccalaureate Organization, 2005; Idaho State Department of Education, n.d.; Juarez-Coca, 2012; Marshall & Andrews, 2002; North & Jacobs, 2010; Oregon Department of Education, 2014; Richardson, 2007; Rubenstein, 2012; Tinberg & Nadeau, 2011; Waits, Setzer, & Lewis, 2005). States are passing laws and policies allocating tax dollars to pay for high school students' participation in dual credit courses (Idaho State Department of Education, n.d.; North & Jacobs, 2010) and national magazines use the enrollment numbers of dual credit courses as a basis for ranking high schools (Moris, 2014). With the current focus on dual credit courses, I felt it prudent to examine the origin, benefits, and drawbacks of dual credit courses, focusing on concurrent enrollment (CE) courses.

After reviewing the literature, I found that little research had been done to study the academic benefits of dual credit courses. A few studies have been done to examine how well students perform after they have completed a dual credit course (Hébert, 2001; Marshal & Andrews, 2002; North & Jacobs, 2010; Rubenstein, 2012); however, little is known about how much students learn while in a dual credit course.

Program Choices

Dual credit courses started in the 1950s as a way to increase student engagement by offering the juniors and seniors a way to take more challenging courses and get rewarded by earning college credit while still in high school (Marshall & Andrews, 2002; Tinberg & Nadeau, 2011; Watt-Malcolm, 2011). The literature describes three basic models for dual credit. Test based credit, such as Advanced Placement (AP) College Board or International Baccalaureate (IB) programs, high-school based courses, such as concurrent enrollment (CE) courses, and college-based courses such as Early College High Schools (ECHS).

Test-based programs such as AP and IB programs provide students with tests they can take in a variety of subjects (College Board AP, 2009; Juarez-Coca, 2012). Students usually take a course designed to prepare them for a specific test (College Board AP, 2009, International Baccalaureate Organization, 2005). However, there are no requirements to take an AP test (College Board AP, 2009). Students taking an AP test receive a score ranging from 1 to 5 (Juarez-Coca, 2012). After the test is scored, the College Board will report the score to as many universities or colleges as the student taking the test would like (College Board AP, 2009; Juarez-Coca, 2012). The score determines the number of college credits earned.

IB programs differ slightly from AP courses in that students must take a qualified course from a trained instructor in order to take the test (International Baccalaureate Organization, 2005). The IB organization qualifies the teachers and places additional restrictions on the IB classroom, such as class size and duration (International Baccalaureate Organization, 2005). As a result, there is an investment of time and money required from both the teacher and school in order for an IB program to become established (International Baccalaureate Organization, 2005). After taking an IB course, students are allowed to take an IB test, and the scores are reported to universities and colleges at the student's request. Again, the score determines the number of college credits earned.

Both AP and IB tests are given nationally and are supposed to be aligned with the published educational standards (Juarez-Coca, 2012). Universities and colleges set their own criteria for accepting credits based on AP and IB test scores and if the student score is high enough, the student will receive credit for a course in the college or university in participating AP or IB programs (College Board AP, 2009; Juarez-Coca, 2012). If a student meets the scoring criterion, a pass/fail grade is usually given to the student in the appropriate course (Juarez-Coca, 2012).

ECHS models select students from a pool of applicants to attend a high school that shares a campus with a community college and students are able to take college courses from a college instructor (Edmunds et al., 2010). With their high school degree, the students are able to earn up to a two year degree from the participating college (Edmunds et al., 2010). Although ECHS programs are effective at educating, engaging, and motivating students, ECHS programs are costly, which prohibits them from becoming widespread (Edmunds et al., 2010).

Concurrent Enrollment (CE) programs enroll students in college or university courses at their local high school and are taught by high school teachers (Juarez-Coca, 2012). The teacher is qualified by the college or university and the curriculum must be preapproved by the college or university where the student receives college credit for taking the course (Juarez-Coca, 2012). At the end of the course, the high school teacher assigns the student a letter grade at both the high school and university and the credits appear on the university transcript as they would had the course been completed at the college or university (Juarez-Coca, 2012).

More high schools are offering dual credit courses than in the past (Farkas & Duffett, 2009; Marshall & Andrews, 2002; North & Jacobs, 2010) perhaps as a result of social pressure put on schools in two forms. First, political leaders have been passing laws and encouraging students to participate in dual credit courses (Idaho State Department of Education, n.d.; North & Jacobs, 2010). Second, private organizations publish national and state-wide school ranking and the organizations use the number of students taking and passing dual credit courses as one of many metrics to determine school ranking (Moris, 2014). Why is there a public interest in increasing the number of students in dual credit courses? What benefits does the public see that drive them to push for more dual credit course participation?

Benefits of Dual Credit Courses

There are two benefits of dual credit courses that regularly appear in the literature. First, students who take dual credit courses are more likely to receive a college degree within four years of graduating from high school (College Board AP, 2009; Edmunds et al., 2010; Marshall & Andrews, 2002). Students who take an AP course are 62% more likely to graduate from college within four years than students who do not take AP courses (College Board AP, 2009). Students who have taken dual credit courses in high school can enter their first year at college as a junior, saving the student two years of college time (Marshall & Andrews, 2002).

The second claim that is consistently reported in the literature is that students who take advantage of dual credit programs save money (College Board AP, 2009; Hébert, 2001; Juarez-Coca, 2012). Generally, students who receive dual credit at their college spend fewer semesters in college and can save on average $26,000 for every year they don't spend in college (College Board, 2009; Marshall & Andrews, 2010).

Not all universities accept credits earned in high school; therefore, to realize the savings described above, a student must attend a university that accepts dual credits. Test based dual credits, such as AP and IB, are the most widely accepted credits (College Board AP, 2009; Juarez-Coca, 2012). CE credits are generally accepted at only the college or university that approved the instructor and curriculum, but it is becoming more common for universities to transfer these credits and CE credits are just like any other university credits (Juarez-Coca, 2012).

Beyond potentially saving students a significant amount of time and money, there are other benefits of dual credit programs described in the literature. First, dual credit programs have re-engaged high school seniors (Tinberg, & Nadeau, 2011). In recent times, many college bound students have stated that their final year in high school was time wasted because it lacked significance (Richardson, 2007). Richardson (2007) claimed that seniors in high school who had already completed the core requirements and didn't have interest in electives found no value in their senior year of high school. With the introduction of dual credit programs, many high school students have found something meaningful in their senior year (Bryant, 2001; Marshall & Andrews, 2002; Tinberg, & Nadeau, 2011).

Taking dual credit courses often helps students get accepted into more exclusive colleges (College Board AP, 2009). Thirty-one percent of selective universities consider students' AP experience, and 85% of university's admittance officers consider AP experience to have a favorable impact on a student's chance to be accepted into the school (College Board AP, 2009). Finally, students taking dual credit courses are nearly twice as likely to earn a college degree as students who do not (Rubenstein, 2012).

Limitation of Dual Credit Courses

While a lot of evidence points towards dual credit courses helping students get into and through college, there is no conclusive evidence that suggests students' learning is improved due to dual credit courses (Hébert, 2001; Tinberg, & Nadeau, 2011; Watt-Malcolm, 2011). While it is true that students who have participated in dual credit courses tend to do better when compared to students who have not taken any dual credit courses, Richardson (2007) argued that the post-secondary success was explained by the selection process. That is, when you only accept the best students into a dual credit course, the participants in the course will, on average, do better than the general population at college.

Bryant (2001) conducted a study that uncovered concerns about the instructional quality of dual credit courses. Bryant claimed that there are typically stricter hiring requirements for the instructors at universities than at high schools, resulting in teachers at the high school level that are not able to provide students with instruction comparable to that of university teachers. Further, the high school teachers may not understand how their class fits into the broader area of study (Bryant, 2001). Bryant (2001) concluded high school instructors were not able to prepare students in introductory courses as well as the instructors at the university. This conclusion may be accurate, but it is important to note that Bryant's study addressed the quality of instruction but did not examine student learning.

Bryant's (2001) study also uncovered a concern that the money being spent on dual credit systems is being spent unfairly. Bryant surveyed residents of Texas about dual credit offerings and the results showed that the participants in the survey felt tax money spent at high schools to support students getting college credits was unfair to taxpayers. My research is not concerned with whether or not spending tax money on college credits for high school students is fair, but states are choosing to spend taxpayers' money on college credits for high school students (Bryant, 2001; Idaho State Department of Education, n.d.; Oregon Department of Education, 2014). If more states are spending more money on dual credit courses, then we can expect that more citizens may question the value of dual credits.

A survey of teachers conducted by Farkus and Duffett (2009) suggests that students are not taking dual credit courses in order to challenge themselves academically, contradicting Tinberg's report from 2011, rather 90% of the teachers surveyed said students were taking the course only to make their college application more appealing to selective schools, and 75% of teachers believed that high schools only push dual credit courses so that their rankings look better. Fifty-six percent of teachers felt students in their class had overestimated their abilities and could not rise to the standards of the class, and 60% said their students were only there because parents wanted them in the class (Farkus & Duffett, 2009). The results of Farkus and Duffett (2009) suggest that the purpose of dual credit courses is moving away from engagement and challenge and moving toward getting students into college.

The data reported by Farkus and Duffett (2009) support the idea that students in high school may not be ready for the standards of university level work and that they may not be motivated enough to meet those standards. As the focus of dual credit courses moves away from getting students through college, it becomes more important to make sure students taking dual credit courses are able to gain the same knowledge as the students at a college or university taking the same course.

Measuring Learning

This investigation compared learning for two groups of students; therefore, in this section, different approaches to measuring learning are described. Two common approaches to measuring learning are constructed-response tests (e.g., essay or short answer tests) and selected-response tests (e.g., multiple-choice tests). The advantages and disadvantages to each approach to testing has been the subject of debate (Falmagne Cosyn, Doignon, Thiéry, 2006; Hunt, 2003; McMillan, 2011). One advantage to constructed-response tests is they can assess deeper levels of understanding and identify misconceptions held by examinees; however, these tests often cannot be reliably scored (Popham, 2014). In contrast, selected response tests can be reliably scored, but often assess surface learning—rather than deeper levels of learning. Additionally, an incorrect answer often tells little about a student's understanding of the concept being tested (Hunt, 2003). However, selected-response tests have some strengths as well (Falmagne et al., 2006; McMillan, 2011). A well written selected-response test can eliminate subjectivity in scoring (McMillan, 2011), which improves the reliability of the test. Further, performance on individual items is objectively derived; therefore, the overall score provides a reliable ranking of students' knowledge (Falmagne et al., 2006). In my research, I compared the knowledge of two groups of students, a high school CE class, and a university class enrolled in the same university class as the high school students.

My primary concern was using a reliable test; therefore, I used a selected-response test. Moreover, to assess deeper levels of learning, I used a special kind of selected-response test.

Concept Inventories

A concept inventory (CI) is a selected-response test designed to assess students' conceptual understanding of a specific topic (Evans et al., 2003; Marbach-Ad et al., 2010; Whitney, 2011). CIs are commonly designed using a 4-phase process that results in a test that can identify students' depth of conceptual thinking. Phases 1 and 2 involve identifying the scope of the test and specific topics to be assessed. These phases are generally completed by a community of professional educators through collaborative conversations and based on best educational practices within a given subject (Adams & Wieman, 2011). Phases 3 and 4 are used to probe for students' naïve conceptions and develop the final form of the selected-response test to be given to students (Adams & Wieman, 2011). Adams and Wieman (2011) outlined the steps involved in Phases 3 and 4 as follows:

> (1) Establish topics that are important to teachers (in our case, college or university faculty members).
>
> (2) Through selected interviews and observations, identify student thinking about these topics and the various ways it can deviate from expert thinking.
>
> (3) Create open-ended survey questions to probe student thinking more broadly in test form.
>
> (4) Create alternatives for the selected-response test that measure student thinking—using the responses generated in Step 3 as distractors.
>
> (5) Carry out validation interviews with both novices and subject experts on the test questions.

(6) Administer to classes and run statistical tests on the results.
Modify items as necessary.

The statistical tests referred to in Step 6 include testing for reliability (Adams & Wieman, 2011). After completing all 4 phases, CIs can be used to accurately probe students' understanding of common phenomena in science and other subjects (Evans et al., 2003; Marbach-Al et al., 2010; Whitney, 2011).

Inventories have already been designed to assess students' knowledge in several areas including (and going beyond) biology, genetics, statistics, and density (Evans et al., 2003; Foundation Coalition, 2008; NC State University, 2007). My study focused on the performance of students in a CE physics course and CIs are good tools to measure students' knowledge in physics (Adams & Wieman, 2011; Evans et al., 2003; Whitney, 2011).

In my study, I used the Force Concept Inventory (FCI) to measure understanding of physics, which is a CI. The initial version of the FCI was developed and validated by Halloun and Hestenes (1985). The FCI and its revision in 1995 are generally accepted as reliable tests (Halloun & Hestenes, 1985; Hestenes, Wells, & Swackhamer, 1992; Huffman & Heller, 1995; Lasry, Rosenfield, Dedic, Dahan, & Reshef, 2011; Savinainen & Scott, 2002a).

Although there was never a large scale study of the 1995 version of the FCI's reliability or validity (Lasry et al., 2011), the issue has been overlooked mainly because much of the content was pulled directly from the original test, and many experts felt validation was unnecessary (Hestenes & Halloun, 1995, Hestenes et al.1992; Huffman & Heller, 1995; Lasry et al., 2011). Additionally, there is a strong correlation between the

FCI and the Force and Motion Concept Evaluation (FMCE), which established the validity of the FCI (Lasry et al., 2011; Savinainen & Scott, 2002).

Gender Bias in the FCI

I have found little evidence in the literature for any bias in the FCI except for gender. It appears that there is a significant difference in the performance on the FCI between males and females, males tend to perform better than females on the FCI (Colletta, Phillips & steinert, 2012; Dietz, Pearson, Semak, & Wills, 2012; McCullough, 2004).

Using the Mantel-Haenszel method of measuring differential item functioning, Dietz et al. (2012) showed that there are a few questions in the FCI that favor one gender over the other. This method of analysis breaks the sample into performance bands and looks for bias within each band. Dietz et al. (2012) showed that there was far less bias than was previously measured. There were only three questions that showed bias with a certainty greater than 99.5% (Dietz et al., 2012). Of the three questions showing bias within the performance bands, two of them favored females and one favored males (Dietz et al., 2012). These data contradicts the previously held belief that the FCI strongly favors males. It needs to be remembered that a much higher percentage of males were sorted into the higher performance strata, with females only making up 10% of the highest strata (Dietz et al., 2012). This study indicates that while men perform better on the FCI overall, females outperform males within their performance bands (Dietz et al., 2012). This indicates that while males score high on the test, and there is bias in the results, the bias is not likely to be from the test. The bias is likely generated elsewhere (Dietz et al., 2012).

McCullough (2004) attempted to correct the apparent male favored gender bias in the FCI by rewriting the questions, replacing more masculine or gender-neutral items with particularly feminine ones. For example, steel bearings rolling off a table became oranges rolling off the kitchen table and pictures of male-looking subjects were replaced with female-looking subjects (McCullough, 2004). When the two tests were compared, it turned out that it had no overall effect on females' scores, and had a negative effect on males' score (McCullough, 2004). McCullough (2004) expected that there would be a significant effect on the females' scores but the results were not consistent with McCullough's hypothesis. McCullough (2004) compared the effect on each of the re-written questions and found both positive and negative effects on individual questions, which balanced out over all 30 questions, indicating that the bias is not inherent in the test, but stems from something else.

Interpreting the FCI

The authors of the FCI intended for it to measure six factors: Kinematics, Newton's fist law, Newton's second law, Newton's third law, the superposition principle, and kinds of force (Huffman & Heller, 1995). There are a total of thirty questions in the FCI that address the different dimensions with seven questions diagnosing Kinematics, nine diagnosing the First Law, five diagnosing the Second Law, four diagnosing the Third Law, five diagnosing the Superposition Principle, and a total of thirteen questions that address different kinds of force (Hestenes et al., 1992; Huffman & Heller, 1995).

The division of the FCI into subgroups has been an issue of debate (Huffman & Heller, 1995; Scott, Schumayer, & Gray, 2012). Huffman and Heller (1995) conducted a factor analysis in which they determined, based on the results of student tests, that the

FCI actually should be separated into ten factors. The correlation of these factors were so weak that Huffman and Heller (1995) questioned the ability of the FCI to measure any coherent concept. A more recent factor analysis, performed by Scott et al. (2012), was able to divide the FCI into just five factors, but again the factors were unreliable. Scott et al (2012) reached the conclusion that students don't break physics into three laws, instead students' understanding of Newtonian mechanics is much more fractured, and as a result the subgroups of FCI should not be analyzed, but rather the FCI should be taken as a single factor testing students' understanding of the Newtonian model of kinematics. There is general consensus that the FCI should be taken as a whole and that it should not be pieced apart in order to assess just one factor of Newtonian mechanics and when the FCI is separated into smaller segments the test loses some of its reliability (Huffman & Heller, 1995; Hestenes & Halloun, 1995; Lasry et al., 2011; Scott et al., 2012).

When interpreting the FCI, a common practice is to use the normalized gain and it is often considered to be a better practice than using just the gain, as the gain often shows a strong negative correlation with the prescore (Hake, 1998). Gain can be measured by finding the difference from the post-instruction assessment score and the pre-instruction assessment score. Normalized gain is the ratio of the gain to the amount of gain that was possible:

$$Normalized\ gain = \frac{Postscore - Prescore}{Total\ possible - Prescore}$$

There are likely influences other than the structure of the class, and the instruction given by the teacher that impact students' normalized gains on the FCI (Colletta & Phillips, 2005; Meltzer, 2002; Nieminen, Savinainen, & Viiri, 2012). It has been shown that there is a strong correlation between students' FCI normalized gain and students' mathematics preparation (Meltzer, 2002), scientific reasoning (Colletta & Phillips, 2005; Nieminen et al., 2012), and representational consistency (Nieminen et al, 2012). When comparing the instructional quality of classes, it becomes important to take students' abilities in these areas into consideration as well as the normalized gain on the FCI (Nieminen et al., 2012).

Gender bias has also been observed in the Lawson test for scientific reasoning, with males out performing females (Colletta et al., 2012). This bias combined with the strong correlation between the normalized gain on the FCI and a student's scientific reasoning ability (Colletta & Phillips, 2005; Meltzer, 2002; Nieminen et al., 2012) may be the reason we see a gender bias in the FCI.

Bruun and Brewe (2013) suggested that although the FCI only measures students' conceptual knowledge of Newtonian mechanics, there is a strong enough correlation between students' performance on the FCI and their overall understanding of physics that the FCI "functions as a measure of the individuals understanding of physics…" (p. 020109-11). Since the FCI can also be used to predict how well students will perform in physics, I will use the FCI to measure students' knowledge in an introductory physics class and examine the subtest scores to determine not only which group learned more but if there was a difference in what the students learned.

Overview of Study

Students who take advantage of dual credit programs can get a head start on their college career, resulting in the students saving both time and money. However, studies have not compared learning in dual credit courses and similar courses offered at colleges and universities. The purpose of this study was to do just this. My research question is as follows: Is student learning, as measured by the FCI, different for dual credit courses compared to a college course covering the same content?

CHAPTER III: METHODOLOGY

I attempted to answer my research question by performing a cross sectional, repeated measures comparison. I compared two groups of students enrolled in a Physics 101 course. One group consisted of students enrolled in a concurrent enrollment (CE) course offered at a high school through a local university. The other group consisted of university students enrolled in the same course at the local university. The FCI was administered before and after instruction about Newtonian mechanics. I collected de-identified data from the pre- and post-instruction assessment. I then analyzed the data statistically to examine differences in performance from pre- to post-instruction and between groups.

Sampling Process

I sampled two populations of students enrolled in a Physics 101 course. The first was high school students enrolled in a CE course through a state university. The second was university students enrolled in the same course at the same university.

To identify participants, I emailed three teachers who were involved in the CE program at a local university, as well as the instructor of Physics 101 at the same local university, explained my study and asked for any data they would be willing to share.

After the teachers had provided the data, I examined all the data to make sure the following qualifications were met:

1. The data came from students who were in a class that allowed them to earn credit in Physics 101 at the local university.

2. The pre- and post-instruction assessment data was paired and de-identified.

3. The pre- and post-instruction FCI were given as a regular part of the classwork.

4. Only students who took both the pre- and post-instruction assessment were included in the data.

The three high school teachers shared their data. One set of de-identified data included both pre- and post-instruction data, but students' data were not paired from pre- to post-instruction, so they could not be used in the analysis. Another high school teacher did not administer a pre-instruction assessment and so these data were not used in the analysis. The third teacher was able to supply me with usable data from one class with 21 students. The university instructor was able to provide usable data from 53 students. (Raw data is included in Appendix B.)

The university is a state university with a student population of a little over 22,000 students. The general population of the university is 54% female, 76% white, 40% part time students and a large portion of non-traditional students. The Physics 101 class was a typical cross section of the university's population. The Physics 101 course was designed as a core course intended for all students, not just those students who were interested in physics, and had no prerequisite math course but used some basic algebra. The university Physics 101 class lasted one semester.

The high school is in a large suburban district with approximately 2,200 students attending the school. The population is predominantly white (85%) and 52% female. About 22% of the school is on free and reduced lunch. The physics class was approximately 70% male, but was otherwise similar to the school's population. The CE course was built into an AP class. As an AP class, students were informed that the course would be more demanding than the average high school course. Also, as an AP course, the CE Physics 101 course was designed to accommodate those students with a personal interest in physics, and required a strong background in mathematics. The CE physics course lasted two semesters.

I conducted a comparison of the syllabi for each class and summarized the topics in Table 1. My conclusion is that both courses were designed to cover the same material, with the high school course requiring a couple of extra topics that were not covered in the university course and were not related to the FCI. The language used to describe the content covered causes me to make some assumptions in order to draw my conclusion. I am comfortable making some of these assumptions, for example "Motion in 1D" (at the high school) is the same as "Linear motion" (at the university). Other assumptions are a little less convincing, such as "Magnetism" and "Circuits" (listed separately at the high school) includes "Induction" (listed at the university).

Table 1.
A comparison of topics covered for each population

High School	University
Motion in 1 D	*Inertia*
Projectile motion	*Linear motion*
Vectors	*Newton's 2nd law*
Newton's Laws	*Newton's 3rd law*
Work and Conservation of Energy	Energy
Circular motion	*Gravity*
Gravity	Vibrations and Waves

Momentum	Sound
Torque Rotational Equilibrium	Resonance
Fluid mechanics	*Electrostatics*
Thermodynamics	Electric Current
Waves and Sound	*Magnetism*
Mirrors and Lenses	Induction
Interference	Light
Static Electricity	The Atom and the Quanta Spectra
Capacitors, and Voltage	Fission/Fusion Radioactivity
Circuits	
Magnetism	
Modern Physics (Includes fission/fusion)	
Quantum mechanics	

Note: *Italics* indicate which subjects are covered by the FCI. **Bold** indicate a topic not covered at the university.

I also conducted a comparison of the labs that were planned for each course and summarized them in Table 2. The comparison showed very little overlap between labs for the two courses, but a similar number of labs addressing concepts from the FCI. (Seven at the high school and nine at the university.)

Table 2.
A comparison of planned lab activities

High School	University
Density lab	*Linear Motion*
Acceleration due to gravity	*Free Fall*
Projectiles lab	*Force Table*
Catapults	*Physics of toys*
Friction Lab	*Gravity*
Conservation of energy lab	Waves
Measurement of tangential velocity in Circular motion	Sound

Conservation of momentum	Resonance
Rotational Equilibrium Lab	*Electrostatics*
Pulley lab	Bulbs
Buoyant force and pressure	*Magnetism*
Determination of absolute zero	Simple Motor
Pendulums	Spectra
Bungee cord project	Radioactivity
Lenses	
Diffraction lab	
Simple Circuits	
Capacitors	
Resistance Lab	
Parallel and Series Circuits	
Magnetic Force Lab	
Electromagnetism Lab	
Calculation of Planck's Constant	

Note: *Italics* indicate that the labs would help develop concepts assessed by the FCI.

The FCI was originally designed, with thirty total questions, to cover six subgroups; Kinematics (seven questions), Newton's First Law (nine questions), Newton's Second Law (five questions), Newton's Third Law (four questions), the Superposition Principle (five questions), and different Kinds of Forces (thirteen questions) (Hestenes & Halloun, 1995). Table 3 compares the FCI subgroups to the topics and labs covered for each population.

Table 3.
Coverage for the FCI

	High School		University	
	Topics	Labs	Topics	Labs
Kinematics (7)	Motion in 1 D, Projectile motion, Vectors	Projectiles lab	Inertia, Linear motion	Linear Motion, Free Fall
First Law (9)	Newton's Laws	None	Inertia, Linear motion	None
Second Law (5)	Newton's Laws	Acceleration due to gravity, Friction Lab	Newton's 2nd law	Physics of toys
			Newton's 3rd law, Gravity	None
Superposition Principle (5)	Newton's Laws, Vectors, Gravity	None	Newton's 3rd law, Gravity	None
Kinds of Forces (13)	Gravity, Torque, Rotational Equilibrium, Static Electricity, Magnetism	Acceleration due to gravity, Friction Lab, Rotational Equilibrium Lab, Buoyant force and pressure, Pulley lab, Magnetic Force Lab, Electromagnetism Lab, Bungee cord project	Newton's 2nd law, Gravity, Electrostatics, Magnetism	Free Fall, Force Table, Physics of toys, Magnetism

Note: The number in parenthesis is number of questions about that topic on the FCI

My conclusion is that both classes were designed to cover the material in the FCI and that each class was designed to teach much more than the concepts covered in the FCI. However, not all topics were covered in the form of a lab activity. It does seem that the topics not covered in lab were the same at both the high school and the university.

It should be noted that both of the syllabi suggested that the material covered in the course may be altered based on teacher discretion and I do not know if the courses were altered from their design.

The university class has two components, the lecture and the lab. The lecture is held once a week for 75 minutes in an auditorium with 100+ students enrolled in the class. The lab is held once a week for 2 hours in smaller groups of no more than 24. The course lasts 16 weeks. As a core course, the university course had no prerequisites and used minimal algebra in the class. If the student passes the course, they receive 4 credits.

The high school course was held every other day for 85 minutes over the course of a full academic calendar. The class meet for a total of 78 periods. The instructor planned to designate 25% of the total class time as lab time. The rest would be designated lecture time. The course was offered with an option to take the AP test at the end, and was therefore perceived to be a more challenging course. (The students intending to take the AP course were required to do additional work outside of the regular class.) There was also a prerequisite of Algebra II, although this prerequisite was only a recommendation to the students. At the high school, there was only one class offered with 28 students.

Both the high school and university students were able to earn 4 credit hours for a survey of physics course. The high school students also received 2 semesters worth of lab science credit at their high school.

Design

I had two groups (high school students in a CE course and university students, and pre- and post-instruction test scores. Therefore, I used a 2 (Course: high school CE versus university course) x 2 (Time: pre- versus post-instruction) design. I used the same 2 x 2 design for each of the 6 subgroups and did a simple statistical comparison of gain and normalized gain scores.

Data Analysis

I used a 2 x 2 analysis of variance to examine differences in performance across the courses and across time for the total score as well as each of the six subgroups. I also concluded a simple effect analysis for the total score, gain and normalized gain.

I also used KR20 (Kuder & Richardson, 1937) to determine the internal consistency reliability of the tests at the various times for each group as well as the internal reliability of each subgroup for each test.

CHAPTER IV: RESULTS

The primary purpose on this study was to examine whether student learning, as measured by the FCI, was different for dual credit courses compared to a college course covering the same content. Before exploring differences between groups and across time, I conducted preliminary analyses to assess the reliability of the FCI for my sample.

Internal Consistency Reliability of the FCI.

I conducted an analysis of internal consistency reliability of the pre- and post-instruction assessments at both the university and high school. I used the KR20 formula (Kuder & Richardson, 1937) to determine the reliability coefficients of each of the four assessments. Reliability coefficients can range from 0 to 1, any coefficient below .50 is considered unacceptable, coefficients above .70 are considered acceptable and above .80 they are considered good (Williams, 2014). The results are displayed in Table 4. Three of the four total scores had good reliability. Only the pre-instruction assessment at the university had poor reliability on the total score. The reliability of the subtests were split with the high school assessments having a reliability in the poor to good range, and the university haveing mostly unacceptable reliability.

Table 4.
Reliability coefficients of each assessment and subgroup.

	Reliability coefficients	
	High School	University
Pre-instruction assessment		
Total Score	**.91**	**.59**
Kinematics	.63	.08
First Law	.79	.37
Second Law	.60	.05
Third Law	.89	.63
Superposition Principle	.63	.01
Kinds of Forces	.81	.41
Post-instruction assessment		
Total Score	**.93**	**.83**
Kinematics	.78	.56
First Law	.73	.66
Second Law	.72	.48
Third Law	.88	.72
Superposition Principle	.70	.40
Kinds of Forces	.88	.60

Comparison of Performance on the FCI

A 2 (Course: high school CE versus university course) x 2 (Time: pre- versus post-instruction) analysis of variance was conducted to examine the effects of course and time on student learning (measured by the FCI). The descriptive statistics are presented in Table 5.

The raw data can be found in Appendix B, and additional descriptive statistics can be found in Appendix C.

Table 5.
Total Score Means by Course and Time

	Pre-instruction assessment score	Post-instruction assessment score
High School	11.43 (1.0)	18.86 (1.3)
University	7.25 (.63)	11.66 (.84)

Note: The total number of questions is thirty. The numbers in parenthesis are the standard error of the mean.

As seen in Table 5, performance on the FCI increased significantly across time, $F(1, 72) = 57.2$, $MSe = 18.5$, $p < .001$, *partial eta squared* = .44. The courses also differed significantly, $F(1, 72) = 24.9$, $MSe = 39.3$, $p < .001$, *partial eta squared* = .26. Moreover, there was a marginally significant Course x Time interaction, $F(1, 72) = 3.66$, $MSe = 18.5$, $p = .06$, *partial eta squared* = .05.

To better understand the interaction, tests of simple effects were conducted. These tests showed that scores on the FCI were significantly higher for students in the high school CE course than for students in the university course at both the pretest [$t(72) = 3.57$, $p = .001$] and the posttest [$t(72) = 4.59$, $p < .001$]. The significant interaction means the change from pretest to posttest was not the same for the two courses. As seen in Table 5, FCI scores increased more from across time for students in the high school CE course than for students in the university course.

Table 6.
Kinematics Score Means by Course and Time

	Pre-instruction assessment score	Post-instruction assessment score
High School	3.10 (0.30)	3.71 (0.40)
University	2.02 (0.19)	2.28 (0.25)

Note: The total number of questions in the kinematics subgroup was seven. The numbers in parenthesis are the standard error of the mean.

As seen in Table 6, performance on the kinematics sub group increased significantly across time, $F(1, 72) = 4.04$, $MSe = 1.5$, $p = 0.048$, *partial eta squared* = .053. The courses also differed significantly, $F(1, 72) = 24.8$, $MSe = 39$, $p < .001$, *partial eta squared* = .26. There was not a significant Course x Time interaction, $F(1, 72) = 0.625$, $MSe = 1.5$, $p = 0.44$, *partial eta squared* = .009.

To better understand the interaction, tests of simple effects were conducted. These tests showed that scores on the kinematics sub group were significantly higher for students in the high school CE course than for students in the university course at both the pretest [$t(72) = 3.27$, $p = 0.017$] and the posttest [$t(72) = 2.85$, $p = 0.007$]. The non-significant interaction means the change from pretest to posttest was about the same. As seen in Table 6, kinematics sub group scores increased more from across time for students in the high school CE course than for students in the university course.

Table 7.
First Law Subgroup Score Means by Course and Time

	Pre-instruction assessment score	Post-instruction assessment score
High School	3.81 (0.40)	6.24 (0.46)
University	2.59 (0.25)	4.23 (0.29)

Note: The total number of questions in the first law subgroup is nine. The numbers in parenthesis are the standard error of the mean.

As seen in Table 6, performance on the kinematics sub group increased significantly across time, $F(1, 72) = 49.9$, $MSe = 2.5$, $p < 0.001$, *partial eta squared* = 0.409. The courses also differed significantly, $F(1, 72) = 1.87$, $MSe = 5.4$, $p < .001$, *partial eta squared* = 0.169. There was not a significant Course x Time interaction, $F(1, 72) = 1.87$, $MSe = 2.5$, $p = 0.18$, *partial eta squared* = 0.025.

Table 8.
Second Law Subgroup Score Means by Course and Time

	Pre-instruction assessment score	Post-instruction assessment score
High School	1.71 (0.23)	2.38 (0.30)
University	0.981 (0.14)	1.64 (0.19)

Note: The total number of questions in the second law subgroup was five. The numbers in parenthesis are the standard error of the mean.

As seen in Table 8, performance on the second law sub group increased significantly across time, $F(1, 72) = 12.8$, $MSe = 1.0$, $p = 0.001$, *partial eta squared* = 0.151. The courses also differed significantly, $F(1, 72) = 8.38$, $MSe = 1.9$, $p = 0.005$, *partial eta squared* = 0.104. There was not a significant Course x Time interaction, $F(1, 72) < 0.001$, $MSe = 1.0$, $p = 0.97$, *partial eta squared* < .001.

Table 9.
Third Law Subgroup Score Means by Course and Time

	Pre-instruction assessment score	Post-instruction assessment score
High School	1.47 (0.29)	3.05 (0.30)
University	1.06 (0.18)	1.68 (0.19)

Note: The total number of questions in the third law subgroup was four. The numbers in parenthesis are the standard error of the mean.

As seen in Table 9, performance on the third law sub group increased significantly across time, $F(1, 72) = 27.2$, $MSe = 1.3$, $p < 0.001$, *partial eta squared* = 0.274. The courses also differed significantly, $F(1, 72) = 10.6$, $MSe = 536$, $p = .002$, *partial eta squared* = 0.129. Moreover, there was a marginally significant Course x Time interaction, $F(1, 72) = 5.09$, $MSe = 1.3$, $p = 0.027$, *partial eta squared* = .066.

To better understand the interaction, tests of simple effects were conducted. These tests showed that scores on the third law sub group were not significantly higher for students in the high school CE course than for students in the university course at the pretest [$t(72) = 1.05$, $p = 0.30$]; however, these tests did show that scores on the third law sub group were significantly higher for students in the high school CE course than for students in the university course at the posttest [$t(72) = 4.04$, $p < 0.001$]. The significant interaction means the change from pretest to posttest was not the same for the two courses. As seen in Table 9, third law sub group scores increased more from across time for students in the high school CE course than for students in the university course.

Table 10.
Superposition Principle Subgroup Score Means by Course and Time

	Pre-instruction assessment score	Post-instruction assessment score
High School	1.52 (0.22)	2.476 (0.30)
University	0.774 (0.14)	1.792 (0.19)

Note: The total number of questions in the superposition principle subgroup was five. The numbers in parenthesis are the standard error of the mean.

As seen in Table 10, performance on the superposition principle sub group increased significantly across time, $F(1, 72) = 29.2$, $MSe = 1.0$, $p <.001$, *partial eta squared* = 0.289. The courses also differed significantly, $F(1, 72) = 8.53$, $MSe =1.8$, $p = .005$, *partial eta squared* = 0.106. There was not a significant Course x Time interaction, $F(1, 72) = 0.033$, $MSe = 1.0$, $p = 0.856$, *partial eta squared* < 0.001.

Table 11.
Kinds of Forces Subgroup Means by Course and Time

	Pre-instruction assessment score	Post-instruction assessment score
High School	4.33 (0.45)	7.76 (0.61)
University	2.65 (0.28)	4.57 (0.38)

Note: The total number of questions in the kinds of forces subgroup was thirteen. The numbers in parenthesis are the standard error of the mean.

As seen in Table 11, performance on the kinds of forces subgroup increased significantly across time, $F(1, 72) = 52.7$, $MSe = 4.1$, $p < .001$, *partial eta squared* = 0.42. The courses also differed significantly, $F(1, 72) = 22.7$, $MSe = 7.9$, $p < .001$, *partial eta squared* = 0.83. Moreover, there was a significant Course x Time interaction, $F(1, 72) = 4.16$, $MSe = 4.1$, $p = .045$, *partial eta squared* = 0.055.

To better understand the interaction, tests of simple effects were conducted. These tests showed that scores on the kinds of forces subgroup were significantly higher for students in the high school CE course than for students in the university course at both the pretest [$t(72) = 2.49$, $p = .020$] and the posttest [$t(72) = 3.77$, $p < .001$]. The significant interaction means the change from pretest to posttest was not the same for the two courses. As seen in Table 11, kinds of forces subgroup scores increased more across time for students in the high school CE course than for students in the university course.

Gain Scores

Gain scores were calculated by finding the difference between the post-instruction assessment and the pre-instructions assessment. Any negative score would indicate that the student received a higher score on the pre-instruction assessment than the post-instruction assessment. Gain scores were compared between groups for the total score, as well as the subtest. The mean gain scores are presented in Table 12. The FCI total gain score were marginally significantly higher for students in the high school CE course than for students in the university course [$t(72) = 1.967$, $p = .056$]. The tests also showed that the high school CE students achieved a significantly high gain score in the kinds of forces subgroup [$t(72) = 2.11$, $p = .042$]. There was also a marginally significantly higher gain score for students in the high school CE course than for students in the university course in the third law subgroup [$t(72) = 2.00$, $p = .055$]. Differences in gain scores were found to be non-significant in the Kinematics [$t(72) = 0.85$, $p = .401$], first law[$t(72) = 1.39$, $p =$

.171], second law [t(72) = 0.017, p = .99], and superposition [t(72) = 2.49, p = .020] subgroups.

Table 12.
Gain score means for FCI and each subgroup

	High School	University
FCI Total	7.43 (1.2)	4.43 (0.85)
Kinematics	0.62 (0.34)	0.26 (0.24)
First Law Subgroup	2.43 (.047)	1.64 (0.31)
Second Law Subgroup	0.67 (0.31)	0.66 (0.20)
Third Law Subgroup	1.57 (0.42)	0.62 (0.20)
Superposition Principle Subgroup	0.95 (0.32)	1.02 (0.19)
Kinds of Forces Subgroup	3.43 (0.59)	1.92 (0.40)

Note: Numbers in parenthesis are the standard error of the mean.

Normalized Gain Scores

Normalized gain scores were found using the following equation:

$$Normalized\ Gains = \frac{Postinstruciton\ Score - Preinstruction\ Score}{Total\ Possible\ Score - Preinstruciton\ Score}$$

Normalized gain scores were compared between groups for the total score, as well as the subtest. The mean normalized gain scores are presented in Table 13. The total normalized gain score were significantly higher for students in the high school CE course than for students in the university course [t(72) = 3.01, p = .005]. The tests also showed that the high school CE students achieved a significantly higher normalized gain score in the kinds of forces subgroup [t(72) = 2.81, p = .006]. There was also a marginally significantly higher normalized gain score for students in the high school CE course than for students in the university course in the first law [t(72) = 1.69, p=.10] and third law [*t*(72) = 2.00, *p* = .055] subgroup. Differences in normalized gain scores were found to be non-significant in the Kinematics [*t*(72) = 1.12, *p* = .28], second law [*t*(72) = 0.303, *p* = .76], and superposition [*t*(72) = .424, *p* = .68] subgroups.

Table 13.
Normalized Gain Score means for FCI and each subgroup

	High School	University
FCI Total	.41 (.07)	.18 (.04)
Kinematics	.14 (.09)	.02 (.05)
First Law Subgroup	.44 (.10)	.24 (.06)
Second Law Subgroup	.16 (.10)	.12 (.05)
Third Law Subgroup	.54 (.20)	.17 (.07)
Superposition Principle Subgroup	.25 (.09)	.22 (.05)
Kinds of Forces Subgroup	.24(.04)	.11 (.03)

Note: Numbers in parenthesis are the standard error of the mean. The sample size varied because a perfect score on the pretest within any category makes it impossible to calculate the normalized gain. For the High School Group, the FCI total, First Law, Second Law, and Kind of Force subtests n = 20. For Kinematics and Superposition Principle subtests n = 19. For the Third Law subtest n = 16 for the High School group and n = 51 for the College group.

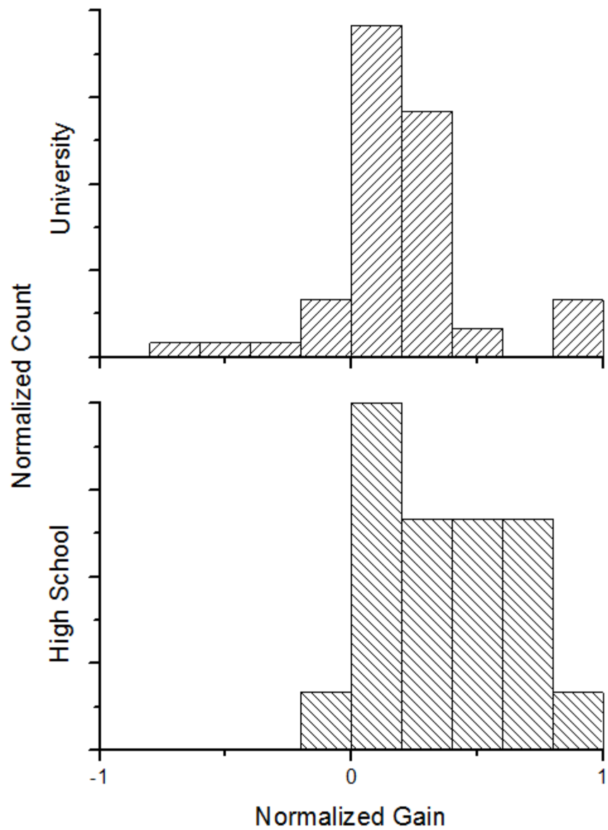

Figure 1. Stacked histogram of normalized gain from the university and high school population. Bin counts have been normalized to increase ease of interpretation. (University *n*=53, High School *n*=20)

Figure 1 shows that the university growth ratios are clustered just above zero with a small number of students demonstrating very high growth. The growth ratios for the high school students seemed clustered around one half, with a much smaller variance.

CHAPTER V: DISCUSSION AND CONCLUSIONS

Discussion

The results of my study suggest the high school students outscored the university students. My statistical analysis suggests that the high school students have a greater conceptual knowledge of Newtonian mechanics going into the course, gain more conceptual knowledge about Newtonian mechanics during the course, and hold greater conceptual knowledge of Newtonian mechanics as they exit the course.

The data from my research does support a positive answer to my research question. CE students enrolled in an introductory physics course were able to acquire the same conceptual knowledge of forces as university students enrolled in the same course. This is supported by both the post-instruction scores and the growth scores. The high school students scored significantly higher on the post-instruction assessment, suggesting that they held greater conceptual knowledge as measured by the FCI than the university students had at the end of the course. The high school students also achieved high gain scores, indicating that the high school students achieved greater gains in knowledge as measured by the FCI.

This is not to say that the course offered at the high school is necessarily better than the university course, or that the instructor is any more skilled. The data only support that the high school students were able to acquire more knowledge in the areas assessed by the FCI.

The poor internal reliability coefficient of the university students' pre-instruction assessment may reduce the certainty of any claims about the scores. Low reliability scores on pre-instruction FCI are not uncommon and have been discussed by Larsy et al. (2011). One conclusion that can be drawn from the low reliability is that students may have a fractured model of Newtonian mechanics and when presented with different scenarios, students may apply different models to reach an answer, which may result in a low reliability coefficient (Lasry, Rosenfield, Dedic, Dahan. & Reshef, 2012). The report by Larsy et al. does not state that we should disregard a low reliability coefficient on the FCI, but Larsy et al. does point out that a low reliability coefficient on the FCI is not very meaningful if the total score is also low, which is the case with the university students' pre-instruction FCI score.

The reliability of the subgroups are considerably low. The low reliability is likely a result of the low number of participants and the low number of questions, however, the low reliability of the subgroup assessments was expected based on the review of literature. These results corroborate what others have found while studying the FCI over the last 20 years, that the reliability of any subgroup for the FCI is generally low (Huffman & Heller, 1995; Hestenes & Halloun 1995, Lasry et al., 2011).

As the FCI can be used as a predictor to determine how students will perform in physics courses outside of mechanics (Bruun & Brewe, 2013; Cabballero et al., 2012), my study suggests that high school students are capable of acquiring the same conceptual knowledge in physics as university students. When the FCI is broken into sub groups, there is no data that contradicts this suggestion. In each subgroup, the high school students achieved higher gains and normalized gains, although the difference was not significant in the majority of subgroups. The lack of significance taken into account with the low reliability scores suggest to me that the subgroup data is not very informative.

When considering both the gains and the normalized gains, my research suggests that high school students enrolled in CE courses are able to learn as much as the university students enrolled in the same university course, and perhaps the high school students can learn more. Further, Figure 1 suggests that the high school class learned more uniformly than the university students. Gains seen in the university students' data were seen mostly in just a few students who showed very large growth, while most students showed very little growth. The gain seen in the high school students' data was evenly spread with the majority of students showing moderate growth. However, there are a few conditions of my study that cause the results to be weaker than they otherwise would be.

The significant gender bias in the FCI (Dietz et al., 2012; McCullough, 2004) may play a role in the results of my study. The university class in my study is approximately 50% male; the high school class is approximately 70% male. Males tend to perform better on the FCI than females, and the greater proportion of males in the high school class may have been enough to create the difference in score that was present in my study. However, without knowing which scores are attributed to which gender, I could not determine if gender played a significant role in my study.

As pointed out by Hébert (2001), typical high school conditions, such as smaller class size, more instructional time, and parental support at home, might make high school a better environment for learning than a college or university. The high school class I studied was much smaller than the university class and had more instructional time. My results suggest that under these conditions the high school students were able to learn more than the university students, supporting Hébert's claim. It would be interesting research to determine whether or not students at a college or university would perform better if the class size were smaller or the instructional time was longer.

Richardson (2007) suggested that the student selection process for dual credit courses often results in selecting students that either perform better or are more motivated to take the course. There may be evidence of this in my study as well. The course was open to all students, but it was advertised as an AP course designed for students who have an inherent interest in physics. The higher scores on the FCI at the beginning of the course may be the result of the high school students' pre-existing interest in physics. The underlying interest from the high school students may also explain why high school students were able to achieve high gains during the course.

With state governments providing funding to students to pay for dual credit fees (Idaho State Department of Education, n.d.; North & Jacobs, 2010), and schools pushing dual credit programs (Farkus & Duffett, 2009; Moris, 2014), I would speculate that there will be an increase in students taking the courses and that the selection process will become less selective, leading to a reduction in the selection effect described be Richardson (2007) and observed in my study. Further research could be done to analyze if and how the selection process is changing as dual credit courses become more popular and to see if there is an effect on student learning.

Limitations

First, my study was limited in scope, sampling, and duration. Most significantly, it used only one test. It would not be prudent to claim that a single assessment is able to determine all students' knowledge. Also, there was no assessment of students' reasoning abilities, and it has been shown that students' reasoning skills have a large impact on the student's FCI scores. Without the data on student's reasoning skills, it is difficult to draw conclusions about the quality of the course. Therefore, a study such as mine, should be only one of many indicators used to determine if students enrolled in an introductory CE course are receiving a comparable education to the students enrolled in the same course at the university.

Second, the FCI is intended to assess students' learning only in the area of Newtonian mechanics. At both the university and high school the students, were expected to learn much more than just mechanics in their classes. While the FCI can be used as a predictor of student performance in other areas of physics, this study did not examine the parts of the curriculum that were not assessed with the FCI.

Third, there are many desired outcomes for any course that is part of the core requirements at a university, one of which is introducing students to the university culture. The FCI does not measure this or any factor other than conceptual knowledge of Newtonian mechanics. My study did not assess whether CE courses help students assimilate into college or university life.

Conclusions

High school students enrolled in a CE course outscored university students enrolled in the same course on the FCI. High school students scored significantly higher on the pre-instruction assessment, post-instruction assessment, and achieved larger gains overall.

My results corroborate the work of others such as Hébert (2001), which suggest that high school instructors are capable of creating an environment in which students are able to learn at the high levels expected from university students. This may dispel the concerns expressed by Bryant (2001) that high school teachers are underprepared to teach students at the university level.

None of the research I did answered the question about long-term retention of knowledge, nor did it examine how students taking introductory courses in physics through a CE program fair at the university in upper division courses that rely on knowledge gained in the introductory course. Further research should be conducted at the university level to determine if the CE courses have done a comparable job in preparing students for upper division work.

The question still remains: Are dual credit courses a benefit to the students enrolled in the courses? I do not feel the question can be answered by the literature I read or the research I conducted. There are still concerns that have not been addressed such as the burden on taxpayers expressed by Bryant (2001) or the lack of motivation in some schools as reported by Farkus and Duffett (2009). However, there are significant benefits for dual credit. First, there is a considerable amount of money that is dedicated to help students enroll in dual credit courses (Idaho State Department of Education, n.s..; North & Jacobs, 2010). Second, dual credit courses help students save time (Edmunds et al., 2010; Marshall & Andrews, 2002) and money (College Board AP, 2009; Hébert, 2001; Juarez-Coca, 2012; Marshall & Andrews, 202) on their college education. Finally, my research adds to the list of benefits by suggesting that high school students may be able to learn just as much as, if not more than, university students in an introductory course.

REFERENCES

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education,* 33(9), 1289-1312. DOI:10.1080/09500693.2010.512369

Bruun, J. & Brewe, E. (2013). Talking and learning physics: Predicting future grades from network measures and Force Concept Inventory pretest scores. *Physicsal Review ST Physics Education Research*, 9(2), 020109-020121

Bryant, A. N. (2001). Eric review: Community college students: Recent finding and trends. *Community College Review, 29*(3), 77-93.

Byrd, S., Elligton, L., Gross, P., Jago, C. & Stern, S. (2007). Advanced Placement and International Baccalaureate: Do They Deserve Gold Star Status? Retrieved from ERIC at http://files.eric.ed.gov/fulltext/ED499005.pdf

Caballero, M. D., Greco, E. F., Murray, E. R., Bujak, K. R., Marr, M. J., Catambone, R., Kohlmyer, M. A. and Schatz, M. F. (2012). Comparing large lecture mechanics curricula using the force concept inventory: A five thousand student study. *American Journal of Physics*, 80(7), 638-644.

College Board AP (2009). *AP® and the cost of college* [Pamphlet]. New York, NY: College Board AP.

Coletta, V. P. & Phillips, J. A. (2005) Interpreting FCI Scores: Normalixed gain, preinstruction scores , and scientific reasoning ability. *American Journal of Physics.* 73(12), 1172-1182.

Coletta, V. P., Phillips, J. A., & Steinert, J. (2012). FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects. *American Institute of Physics Conference Proceedings*, 2/1/2012, 1413(1), 23-26

Dedic, H., Rosenfield, S. & Lasry, N. (2010). Are all wrong FCI answers equivalent? *American Institute of Physics Conference Proceedings*, 10/25/2010, 1289 (1), 125-128.

Dietz, R.D., Pearson, R. H., Semak, M. R. & Wills, C. W. (2012). Gender bias in the force concept inventory? *American Institute of Physics Conference Proceedings:* V*ol. 1413* (pp.171-174)

Edmunds, J., Bernstein, L., Glennie, E., Willse, J., Arshavsky, N., Unlu, F., et al. (2010) Preparing students for college: The implementation and impact of the early college high school model. *Peabody Journal of Education,85*(3), 348-364.

Evans, D. L., Gray, G. L., Krause, S., Martin, J., Midkiff, C., Notaros, B. M., … Wage, K. (2003, November 5-8). Proceedings from 33rd ASEE/IEEE Frontiers in Education Conference: *Progress on concept inventory assessment tools.* Bolder, CO.

Falmagne, J., Cosyn, E., Doignon, J., Thiéry, N. (2006) The assessment of knowledge, in theory and in practice. *Lecture Notes in Computer Science,* 3874, 61-79.

Farkas, S. & Duffett, A. (2009, April). Growing pains in the advanced placmement program: Do though trade-offs lie ahead? Retrieved from The Thomas B. Fordham Institute at http://edex.s3-us-west-2.amazonaws.com/publication/pdfs/AP_Report_7.pdf

Foundation Coalition, (2008, July 21). *Concept inventories.* Retrieved from http://www.foundationcoalition.org/home/keycomponents/concept

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics course. *American Journal of Physics*, 66(1), 64-74

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

Hébert, L. (2001). A comparison of learning outcomes for dual-enrollment mathematics students taught by high school teachers versus college faculty. *Community College Review. 29*(3), 22-38.

Hestenes, D. & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *The Physics Teacher,* 33(8), 502-506.

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *The physics Teacher,* 30(3), 141-158.

Huffman, D.& Heller, P. (1995). What does the force concept inventory actually measure? *The Physics Teacher,* 33, 138-143.

Hunt, D. P. (2003). The concept of knowledge and how to measure it. *Journal of Intellectual Capital,* 4(1), 100-113. DOI 10.1108/14691930310455414

International Baccalaureate Organization. (2005). Applying for IBCC. *IB Academy.* Retrieved October 28, 2012, from http://www.ibo.org/ibcc/apply

Idaho State Department of Education. (n.d.). New Program For All Idaho Students. Retrieved from https://www.sde.idaho.gov/site/advancedOpp

Juarez-Coca, F. (2012). *Idaho dual credit program* [Brochure]. Boise, ID: Boise State University.

Kuder, G. F. & Richardson, M.W. (1937). The theory of the estimation of test reliability. *Psychometrika,* 2(3), 151-160

Lasry, N., Rosenfield, S., Dedic, A., Dahan, A. & Reshef, O. (2011). The puzzling reliability of the force concept inventory. *American Journal of Physics,* 79(9), 909-912.

Lasry, N., Rosenfield, S., Dedic, A., Dahan, A. & Reshef, O. (2012). Reply to “comment on ‘the puzzling reliability of the force concept inventory,’ by N. Lasry, S. Rosenfield, H. Dedic, A. Dahan, and O. Reshef [Am. J. Phys. 79, 909-912 (2011)]”. *American Journal of Physics,* 80(4), 349-350.

Marbach-Ad, G., McAdams, K., Benson, S., Briken, V., Cathcart, L., Chase, M., … Smith, A. C. (2010). A model for using a concept inventory as a tool for students’ assessment and faculty professional development. *CBE Life Sciences Education,* 9(4), 408-416. doi: 10.1187/cbe.10-05-0069

Marshall, R. P. & Andrews, H. A. (2002). Dual-credit outcome: A second visit. *Community College Journal of Research and Practice*, 26(3), 237-242.

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

McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed.). Boston, MA: Pearson Education Inc.

Meltzer, D.E. (2002). The relationship between mathematics preparation and conceptual learning gains in physics: A possible ‘‘hidden variable’’ in diagnostic pretest scores. *American Journal of Physics.* 70(12), 1259-1268.

Moris, R. (2014, April 21). How U.S. News Calculated the 2014 Best High Schools Rankings. Retrieved on June 4, 2014, from http://www.usnews.com/education/best-high-schools/articles/2014/04/21/how-us-news-calculated-the-2014-best-high-schools-rankings

NC State University, (2007). Assessment instrument information page. Retrieved on 7/5/2014 from http://www.ncsu.edu/per/TestInfo.html

Nieminen, P., Savinainen, A., and Viiri, Jouni (2012). Relations between representational consistency, conceptual understanding of the force concept and scientific reasoning. *American Physical Society*, 8(1), 010123-010132

North, T. & Jacobs, J. (2010). Dual Credit in Oregon 2010 Follow-up: An Analysis of Students Taking Dual Credit in High School in 2007-08 with Subsequent Performance in College. Retrieved from www.ous.edu/dept/ir/reports

Oregon Department of Education (2014) Career and Technical Education. Retrieved from http://www.ode.state.or.us/search/results/?id=222

Popham, W. J. (2014). *Classroom assessment: What teachers need to know*. Boston: Pearson.

Richardson, T. (2007). Dual-credit: “A key to the future.” *On the Horizon. 15*(4), 239-244. doi: 10.1108/10748120710836255

Rubenstein, J. (2012). Study: “Dual enrollment” students more likely to attend, graduate from college. *Jobs for the Future.* [Web log post] Retrieved October 28, 2012, from http://www.jff.org/media/news-releases/2012/study-%E2%80%9Cdual-enrollment%E2%80%9D-students-more-li/1476

Savinainen, A. & Scott, P. (2002). The force concept inventory: A tool for monitoring student learning. *Physics Education,* 37(1), 45-52

Scott, T. F., Schumayer, D., & Gray, A. R. (July 01, 2012). Exploratory factor analysis of a Force Concept Inventory data set. Physical Review Special Topics - Physics Education Research, 8, 2.) Steinberg, R. N. & Sabella, M. S. (1997) Performance on multiple-choice diagnostics and complementary exam problems. *The Physics Teacher*, 35 (3), 150-155.

Tinberg, H. & Nadeau, J. (2011). Contesting the space between high school and college in the era of dual-enrollment. *College Composition and Communication, 62*(4), 704-725.

Waits, T., Setzer, J.C. & Lewis, L. (2005) *Dual credit and exam based courses in U.S. public high schools: 2002-03* (NCES 2005-009). U.S. Department of Educaiton. Washington, DC: National Center for Education Statistics.

Watt-Malcolm, B. (2011). Dual credit: Creating career and work possibilities for Canadian youth. *Canadian Journal of Education. 43*(2), 256-276.

Whitney, H. M. (2011, February 9). Concept inventories: Beyond physics. Retrieved from http://chronicle.com/blogs/profhacker/concept-inventories-beyond-physics/30471

Williams, Y. (2014). Test-retest reliability coefficient: examples, lesson & quiz. Retrieved from http://education-portal.com/academy/lesson/test-retest-reliability-coefficient-examples-lesson-quiz.html#lesson

APPENDIX A

The FCI and Its Contents

The content of the FCI would be included in this paper but due to the extensive use and the need for the assessment questions to be novel, the research group from Arizona State University currently in charge of safe guarding the FCI have asked that the inventory itself not be included in any writings about it. However, one can find the FCI on the internet and directions for accessing its contents at the following website: http://modeling.asu.edu/R&E/Research.html

For further information about the FCI visit the following website: http://modeling.asu.edu/R&E/Research.html

APPENDIX B

Institutional Review Board Letter of Approval

BOISE STATE UNIVERSITY

RESEARCH AND ECONOMIC DEVELOPMENT
Research Compliance

August 13, 2014

Mr. Eric Thies (Master's Student)
Department of Curriculum, Instruction and Foundational Studies
1910 University Drive
Boise, Idaho 83725-1745

Mr. Thies,

On October 21, 2013, the Office of Research Compliance, in agreement with the chair of the Social & Behavioral Institutional Review Board (IRB), withdrew your protocol application titled, "Comparison of Physics 101 CE and Regular Students Using the FCI."

It was determined that your study did not meet our definition of "human subjects research." Your study involved the analysis of de-identified or coded data. Because your study did not involve any interaction or intervention with participants or access to private identifiable data, your research was determined not to involve "human subjects," and, therefore, no IRB review or determination of exemption was required.

Thank you for initially submitting a protocol application to this office so we could determine that an IRB review was not needed. Please let me know if you have any questions or need any additional information from me.

Thank you,
Amy J. Smith
IRB Coordinator

APPENDIX C

DE Identified Student Scores on the Pre- and Post-Instruction FCI

CE Student FCI Scores

Student #	Pre-Instruction Assessment	Post-Instruction Assessment
1	3	5
2	4	7
3	5	18
4	6	13
5	6	24
6	7	13
7	7	18
8	8	23
9	9	20
10	9	25
11	10	9
12	10	14
13	10	15
14	12	25
15	13	23
16	15	17
17	15	17
18	17	30
19	20	21
20	24	29
21	30	30

University Student FCI Scores

Student #	Pre-Instruction Assessment	Post-Instruction Assessment
1	1	5
2	2	9
3	2	25
4	3	6
5	3	8
6	3	27
7	4	5
8	4	7
9	4	7
10	4	13
11	5	17
12	5	12
13	5	7
14	5	14
15	5	10
16	5	5
17	5	14
18	5	9
19	6	13
20	6	4

21	6	11
22	6	12
23	6	5
24	6	18
25	6	14
26	6	11
27	7	15
28	7	10
29	7	15
30	7	12
31	8	11
32	8	11
33	8	11
34	9	16
35	9	9
36	9	14
37	9	9
38	9	10
39	9	16
40	9	10
41	9	10
42	10	13
43	10	27
44	10	12

45	10	12
46	10	12
47	11	10
48	12	10
49	12	7
50	12	13
51	12	4
52	15	3
53	17	28