UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

CONCEPT QUESTIONS WITH PEER-LED VS. TEACHER-LED DISCUSSIONS IN A HIGH SCHOOL PHYSICS COURSE

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF EDUCATION

By

KEVIN WARREN
Norman, Oklahoma
2011

CONCEPT QUESTIONS WITH PEER-LED VS. TEACHER-LED DISCUSSIONS IN A HIGH SCHOOL PHYSICS COURSE

A THESIS APPROVED FOR THE
DEPARTMENT OF INSTRUCTIONAL LEADERSHIP
AND ACADEMIC CURRICULUM

BY

[redacted]

[redacted]

[redacted]
Dr. Bruce Mason

Table of Contents

List of Tables

List of Figures

Abstract

Introductory physics classes attempt to guide students while they develop the mental structures and schema necessary to consistently apply the Newtonian concept of force to various situations. Science education research has supported the effectiveness of teaching methods, like Peer Instruction, that require the students to be active participants. This study compared the effectiveness of Peer Instruction when two different types of discussions, peer-led and teacher-led, were used to discuss concept questions in a high school physics class. The study addressed three research questions: (a) does Peer Instruction with peer-led discussions have a positive significant effect on learning (b) does Peer Instruction with teacher-led discussions have a positive significant effect on learning (c) is there a significant difference in learning gains between peer-led and teacher-led discussions. Normalized gain scores between the Force Concept Inventory (FCI) pretest and posttest and the normalized gain scores between concept questions and benchmark tests supported the hypotheses that there was a significant increase between pretest and posttest scores for both groups, but the repeated measures test found no significant difference between the two treatment groups. An important limitation to the study was that there was no significant correlation between the pretest and posttest scores, which limited the ability of the study to support the hypotheses that the specific discussion format caused the change between pretest and posttest scores. Ways to improve the research design to increase the correlation between pretest and posttest scores for future research is discussed.

Chapter I-Introduction

Research Problem

Research about science education in the U.S. shows the need for reforms in education. In 2006, the Program for International Student Assessment (PISA) reported that American 15 year-olds scored below the previous rank on the 2000 assessment in math and science (National Science Board, 2010). The Organization for Economic Cooperation and Development (OECD) administers the PISA assessment every three years, and the U.S. science scores did increase from the 2006 scores. The U.S. scores had been lower than the OECD average, but the 2009 scores were not measurably different than the average. Twelve countries scored higher, twelve countries did not score measurably different, and nine countries scored lower than the U.S. on the 2009 PISA science literacy section (National Center for Educational Statistics, 2011). In 2009, only 60% of twelfth-graders performed at or above the basic level; only 21% of twelfth-graders performed at or above the proficient level; and only 1% of twelfth graders performed at the advanced level on the National Assessment of Educational Progress (NAEP) test for science (National Center for Education Statistics, 2011). The Science, Technology, Engineering, and Mathematics (STEM) Education Coalition is committed to keeping the U.S. the technology leader in the 21st century (STEM Education Coalition, 2011). However, due to poor preparation in high school science classes and further exacerbation by the weed-out mentality of introductory science classes at universities, over 50% of students intending to pursue careers in the natural sciences change their major (DeHaan, 2005).

Poor preparation in science at the high school level could be attributed to teaching methods based on class lectures. These methods have come under attack over the last three decades (Iverson, Briggs, Ruiz-Primo, Talbot III, & Shepard, 2009). Students enter into science classes with non-Newtonian concepts about the world built from their prior experiences, which could include conceptions formed or strengthened during previous science classes (Duit & Treagust, 2003). The problem is that many students leave their introductory science classes with those same non-Newtonian concepts intact (Dilber, Karaman, & Duzgan, 2009). For this study, a Newtonian concept will refer to any idea consistent with the classical view of motion and forces, and a non-Newtonian concept will refer to any view that is not consistent with classical mechanics. Research (Brumby, 1984; Halloun & Hestenes, 1985; Mazur, 1997; McDermott, 1984; West & Pines, 1985) indicates that it is very difficult to replace long-held preconceptions that seem intuitive to the student.

The discipline of science is a classic example of the variance between students' conceptual understandings. The majority of students leave their introductory physics classes frustrated and confused because they have a hodge-podge of two systems – their own, which is ripe with non-Newtonian concepts, and the one presented to them throughout the course – stitched together to explain the world (Mazur, 1997). During physics class, students learn to analyze a small collection of problems with a set of equations, and they view everything else through the lens of their prior non-Newtonian concepts. While students learn accurate concepts, they seldom transfer those concepts into generalized principles to truly understand physics in a way that overcomes their initial non-Newtonian concepts.

To address the problem of non-Newtonian concepts, the process of learning must be reevaluated. According to Crouch and Mazur (2001), physics education needs to move away from the conventional lecture format that has had little affect on student understanding of key concepts. In addition, Mazur (1997) worries that problems from the textbook used in class and for homework assignments emphasize computational rather than analytical problem solving skills.

Need for Research

Many science educators ask the central question, what is the purpose of science education? If a student memorizes a set list of facts about biology, chemistry, and physics, then is that student proficient at science? To some scholars (Marek & Cavallo, 1997), the nature of science is more than just an understanding of a collection of facts; it is viewed as the process of attaining and organizing those facts into a consistent system (Maier & Marek, 2005). When regarded from this perspective, science education can be used as a tool in schools to help students develop higher order rational powers like analyzing, synthesizing, and evaluating (Marek & Cavallo, 1997). A student may never practically need to know how many protons are in oxygen, but the process of learning about the elements can provide opportunities for all students to be challenged to classify the elements, deduce how many subatomic particles are present, and imagine what an atom looks like. Constructing specific science knowledge becomes part of the bigger process of helping students develop the ability to think, which is the central purpose of science education (Educational Policies Commission, 1961).

Physics provides an excellent opportunity to guide students to develop their reasoning abilities and construct schema for new concepts as they move from concrete to formal thinkers. However, if physics education is going to be used to help students develop rational powers, teachers need to find effective ways to present the material that helps students work through their non-Newtonian concepts. To this end, physics education researchers have found that learning strategies can be put through rigorous experimental testing to find which strategies are effective at reducing non-Newtonian concepts (as cited in McDermott, 1990).

Peer Instruction (PI) is one effective method that has been found to reduce non-Newtonian concepts (Mazur, 1997). PI uses concept questions, immediate feedback, and peer-led discussions in conjunction to reduce non-Newtonian concepts. PI was developed to engage students in the learning process, while confronting students' common non-Newtonian concepts. Physics classes where PI was implemented showed significant positive results compared to physics classes taught in the traditional fashion (Crouch & Mazur, 2001; Cummings & Roberts, 2008; Lasry, Mazur, & Watkins, 2008). However, Mazur has not isolated the effectiveness of peer-led discussions over other methods because peer-led discussion is just one component of PI.

Other researchers have also studied the effects of PI. Martyn (2007) focused on comparing two different active learning methods: clickers versus class discussion. No statistically significant difference between the two approaches was found in the study. Kalman, Miner-Bolotin, and Antimirova (2010) found that collaborative group projects were more effective than PI, but that collaborative group projects were also much more time intensive. Nicol and Boyle (2003) found that PI was perceived to be more effective than class-wide discussions by students in large university classes, but no quantitative data were collected in the study.

More research needs to be done to see which methods utilizing PI are more effective in different settings. It is important to collect data to show how peer-led group discussions compare to teacher-led class discussions because class time is important. Teachers who want to implement PI need to be able to justify its effectiveness in the classroom to the school district, other teachers, parents, and students. Additional research could provide more data, which could help to explain how teachers can assist students meaningfully learn the concepts, and to identify the most effective methods in facilitating conceptual change.

Research Question

PI goes beyond the use of clickers as tools for students to answer questions. PI gauges the progress of the class, and in the appropriate places, requires students to reason through the conceptual questions in group settings. Researchers have documented the increase in students' learning gains when using clickers in the classroom compared to a conventional lecture (Auras & Bix, 2007; Martyn, 2007; Wood, 2004; Yourstone, Kraye, & Albaum, 2008). The immediate feedback and higher participation from using clickers could be the main contributing factor leading to the higher learning gains of PI rather than the peer-led discussions. Mazur has not isolated the effectiveness of peer-led discussions over other methods because peer-led discussion is just one component of PI.

The goal of this study was to compare two different active learning methods within PI, the traditional peer-led discussion model and the untested teacher-led discussion. The independent variable for this research study was the discussion type, peer-led and teacher-led. The dependent variable was the learning gain between pretest scores and posttest scores on conceptual understandings of physics concepts, while gender, GPA, and the students current math class served as predictor variables for the study.

There were three main research questions that the study addressed. The first two questions asked if each of the two discussion types would create statistically significant learning gains as evidenced by higher scores on posttests than pretests.

1. Did using peer-led discussions with concept questions create a statistically significant learning gain between pretest and posttest scores?
2. Did using teacher-led discussions with concept questions create a statistically significant learning gain between pretest and posttest scores?

The final question of the study compared the effectiveness of the two discussion types. The purpose of the study was to determine if there was a statistically significant difference between the two methods in the classroom.

3. Was there a measurable difference in learning gains on the FCI pretest and posttest scores between peer-led group discussions versus teacher-led class discussions?

Chapter II-Literature Review

Addressing Non-Newtonian Concepts in Physics

Since the 1970's, researchers have studied preconceptions students hold when they enter science classes (Duit, 2002). These preconceptions about how the world works are not supported by science (Duit & Treagust, 2003). Dilber, Karaman, and Duzgan (2009) have documented that non-Newtonian concepts are common in every area of introductory physics classes. Furthermore, non-Newtonian concepts are a problem at the middle school, high school, and university levels. Studies have shown that students hold misconceptions in mechanics (Clement, 1982; Eryilmaz, 2002; Minstrell, 1982; Towbridge & McDermott, 1980, 1981; Vionnet, 1979), electricity (Basar & Geban, 2008; Cohen, Eylon, & Ganiel, 1983; Dupin & Johsua, 1987; Fredette & Lochead, 1980; Heller & Finley, 1992; Idar & Ganiel, 1985; Maloney, O'Kuma, & Hieggelke, 2001; Sencar & Eryilmaz, 2004), optics (Feher & Rice, 1992; Goldberg & McDermott, 1986, 1987), and thermodynamics (Athee, 1993; Bar & Travis, 1991; Ericson, 1979; Ma-Naim, Bar, & Zinn, 2002; Shayer & Wyllam, 1981).

Even though students are able to learn methods for solving quantitative problems, physics courses taught through lecture are not effective at reducing non-Newtonian ideas about fundamental concepts (Hake, 1998; Halloun & Hestenes, 1985; McDermott, 1990). Educators need to understand how students learn new concepts. In the 1960's, Kuhn (1970) argued that students in science did not progress by just learning more facts, but that students had what came to be known as paradigm shifts that changed how they viewed the information they were presented. The classical conceptual change approach proposed by Posner, Strike, Hewson, and Gertzog (1982) indicates that four conditions must be present before conceptual change will take place within students. First, students need to be given experiences that reveal the inadequacy of their current ideas so that they feel the need for a new model. Second, the new concept must be able to clearly be applied to the new experience to help the students feel they have a better understanding of the world. Third, the new concept must be viewed to have a real potential to be used to reinterpret all of the other experiences that the old concept explained. Fourth, the new concept must also spur new questions for the student (as cited in Dilber, Karaman, & Duzgun, 2009).

Educators are still trying to explain how conceptual change takes place in learners. Bybee and Sund (1990) relate Piaget's theory of the process:

> Intellectual development is an adaptation in response to a discrepancy between the existing cognitive structure and a cognitive referent in the environment. The discrepancy results in a disequilibrium, which produces a reconstruction that brings the system back to equilibrium. (p. 195)

In the 1980's, Piaget's mental functioning model began to be used to understand how students learned new concepts. The mental functioning model consists of assimilation, disequilibrium, accommodation, and organization. Assimilation is the process of an individual beginning to think about the information they are receiving with respect to their current mental structures. If the information does not fit within their current mental structures for understanding the world, then the individual moves into disequilibrium. When in a state of disequilibration, the individual will choose one of several options. Students will either make the information fit their current mental structures or create a new mental structure to accommodate the new information, which is the third step of the mental functioning model. (Another option, although undesired, is when the individual becomes frustrated and then refrains from learning the new information.) The final stage of the mental functioning model is the organization of the new mental structure with all of the other existing mental structures (Marek, 2009).

The key to developing new mental structures is linked to recognizing that current mental structures cannot explain information from a new situation. According to Piaget's mental functioning model, it is necessary for every student to experience disequilibrium in the learning process before equilibration can take place (Bybee & Sund, 1990). Physics researchers should therefore attempt to find developmentally appropriate strategies to provide experiences that create disequilibrium to challenge students' non-Newtonian concepts.

Physics education research (Fagen, Crouch, & Mazur, 2002; Smith, Wood, Krauter, & Knight, 2011) has explored strategies for effectively reducing the number of non-Newtonian concepts for students in the classroom. The problem is that different students will be at different developmental levels. This means that within one classroom, students will understand the concept at different times, which is why constructivist strategies are so vital in the classroom. Constructivist approaches help the teacher and the student communicate more effectively (Powell & Cody, 2009). Teachers need to ask questions to gauge the students' comprehension process and to help avoid the development of new non-Newtonian concepts; however, no one but the student can construct knowledge through assimilation and accommodation. In the constructivist model, students are not passive learners (Gilbert, Osborne, & Fensham, 1982). Students are no longer seen as empty vessels into which teachers pour knowledge. Students come into the classroom with their own ideas about how the world works; therefore, students need to be active participants in constructing their own knowledge (Johnson & Johnson, 1991).

In the 1990's, limitations in the cognitive constructivist approach supported by Piaget's mental functioning model led to the inclusion of the social constructivist approach supported by Vygotsky's learning model (Duit & Treagust, 2003). This acceptance of multi-perspective epistemological frameworks was needed to adequately address the complex process of learning (Duit & Treagust, 2003). Both learning models encourage inquiry methods because they are both based on the idea that students build concepts on existing knowledge that is relevant and meaningful to them. A major difference between the theories is that Piaget believed that thinking led to language, while Vygotsky believed that language led to thinking (Powell & Cody, 2009).

Vygotsky's social constructivism model is driven by the idea that interactions drive knowledge construction. Teachers can provide scaffolding by analyzing and reviewing the concept with the student. The zone of proximal development is the range in which a more capable individual can help guide a person who is learning a new concept to understand the ideas more quickly. It represents the gap between what a person can do independently and what a person is capable of doing with the appropriate assistance from other people. Individuals still construct their own knowledge, but cooperative learning is very important to the theory because social interactions can help the individual construct their understanding (Powell & Cody, 2009).

Peer Instruction

Peer Instruction (PI) applies the research in physics education on conceptual change by laying out the practical steps needed to bring the research into the classroom. PI fits well with Bybee and Sund's (1990) suggestions for implementing educational strategies based on Piaget's mental functioning model. Eric Mazur developed PI in the early 1990's after realizing that students in his physics classes at Harvard University could learn how to perform the mathematical calculations for the problems on his test without understanding the concepts behind the problems. He began to ask conceptual questions in class and on exams to encourage students to think about the problems rather than merely focusing on plugging numbers into an equation to get an answer (Mazur, 1997).

The three components of Peer Instruction.

The three major components of PI include (a) challenging concept questions, (b) immediate feedback, and (c) peer-led discussion groups. PI is built around conceptual questions that cause students with common non-Newtonian concepts to experience disequilibrium (Mazur, 1997). The use of clickers or flashcards encourages student participation and provides immediate feedback to students (Lasry, 2008). Student discussions encourage students to explore the validity of different options and how they fit together.

Concept questions.

Many university researchers have been attempting to make larger classes more interactive by asking concept questions within their lectures (Bligh, 2000; Edwards, Smith, & Webb, 2001; MacGregor, Cooper, Smith, & Robinson, 2000). Mazur started developing PI when he tested his own students and found that they performed better on higher level quantitative questions than on lower level conceptual questions (Mazur, 1997). He argues that textbook questions covered in class and on homework require more mathematical than analytical skills. Nicol and Boyle (2003) found that the majority of students in a PI study reported preferring to answer questions individually before the discussions. Concept questions require students to reason through problems, which helps them develop a deeper understanding of the system, which can then be generalized to broader applications than just memorizing the formula for one specific application.

One key to asking concept questions in class is to include concept questions on the exams (Mazur, 1997). Students need to buy into the importance of understanding the concepts and the context of the problem. If only quantitative problems are asked on the exams, then only problem solving skills need to be learned. Mazur sets the stage for the importance of concept questions in his classes by having a class discussion about the importance of understanding the concepts and by handing out past exams to show the level of understanding needed to answer the conceptual questions on the exam.

Immediate feedback.

Mazur uses clickers in PI to provide instant feedback to the instructor and to hold the students accountable for the information before the exams (Mazur, 1997). Lasry (2008) found that there was no significant difference between using flashcards or clickers to engage students in forming a response and providing immediate feedback in PI. Providing a set of flashcards for each student allows the instructor to quickly survey that students have made a choice by picking up a specific card and assess where the class is at overall on the multiple choice question, but clickers do instantly provide more precise records of responses. This study will use clickers to provide immediate feedback, which is why clickers will be the focus of the remainder of this section.

Clickers are also known as Audience Response Systems, Personal Response Systems, Electronic Voting Systems, Classroom Communication Systems, and Classroom Response Systems. The proliferation of clickers in the classroom is a recent phenomenon. Electronic student response systems began in the 1960's, when Harvard University and Rice University hard wired response systems to desks in some lecture halls. Classtalk, funded by the National Science Foundation, was introduced in 1985, and it revolutionized the market by using graphing calculators for communication platforms (Adams & Howard, 2009). By the 1990's, other companies such as eInstruction and Educue introduced systems that were easier to learn and operate and less expensive. These clickers used hand-held remotes with infrared signals, which allowed the systems to be moved between classrooms. In the last few years, the new trend in universities has been to create web-based systems that only require the addition of software to a laptop rather than the purchase of a hand held remote (Adams & Howard, 2009).

Technology can be an important tool in the classroom, and clickers represent one such technology that has been successfully integrated by educators. Beatty (2004) gives the following description of clickers:

> A classroom response system is technology that allows students to present a question or problem to the class, allows students to enter their

> answers into some kind of device, and instantly aggregates and summarizes students' answers for the instructor. (p. 2)

Clickers allow the teacher to instantly assess student comprehension of the concept, which then allows the teacher to generate appropriately difficult follow-up questions. No one is claiming that putting a box of clickers into a room with students will cause learning to happen. However, many programs have been developed that effectively use clickers to generate immediate feedback and solicit more participation within the classroom (Auras & Bix, 2007; Martyn, 2007; Wood, 2004; Yourstone, Kraye, & Albaum, 2008). Yourstone, Kraye, and Albaum (2008) found that classes given immediate feedback through clickers in class quizzes had greater learning gains than classes that were given the same questions but did not receive the answers until the next class.

Many studies on the use of clickers have also documented that students report liking to use them (Draper & Brown, 2004; Duncan, 2005; Hatch, Jensen, & Moore, 2005; Latessa & Mouw, 2005; Wit, 2003; Zahorik, 1996). In addition, Keller et al. (2007) showed that students have a more positive attitude about using clickers if faculty encourage discussion and are able to get a large part of the class involved in small group discussions. It is also important to note when clickers should not be used. According to Adams and Howard (2009), clickers should not be used to merely have students enter their attendance for the day or to ask high-stakes conceptual questions. Using clickers for attendance makes students feel they are not benefitting from them, and they resent being forced to buy them. Asking high-stakes questions can discourage participation because it can cause students to be more concerned with getting the right answer than exploring the different options (Willoughby & Gustafson, 2009). With high stakes clicker systems low achieving academic students just passively agree with the high achieving academic students who monopolize the discussion when their grade involves getting the correct answer from the discussion (Willoughby & Gustafson, 2009).

Peer-led discussions.

Students must actively engage in the learning process to develop an understanding of the many interconnected concepts in the physics classroom. Student discussions have been shown to be an excellent tool to engage students (Hake, 1998; Halloun & Hestenes, 1985; McDermott, 1984). Schwartz and Bransford (1998) found that having students choose an answer to a concept question before discussing it helped students get more from the discussion that followed the question. These learning gains could be attributed to the idea that student discussions give everyone the opportunity to justify the reasons for their choice and provide alternative reasons. Singh (2005) found that students working in pairs showed greater learning gains together than individually. Aryal and Zollman (2007) also found significant improvement in the students' ability to complete tasks when they worked in groups rather than individually. The groups help students transfer the concepts they previously learned to new contexts within problems. Nicol and Boyle (2003) reported that student responses to interview questions mirrored Vygotsky's idea of scaffolding from peers. Students in PI classes indicated that it was easier to understand concepts from other students who had just mastered the concept than from a teacher.

Grading incentive is also an important part of peer discussions. Fagen, Crouch, and Mazur (2002) support giving a participation grade for concept questions to encourage full participation in the discussions. However, high-stakes grading is discouraged. Willoughby and Gustafson (2009) found that high-stakes grading discouraged group discussions because the stronger academic students became more dominant in the discussion, which allowed the lower achieving students to passively vote with the group. James (2006) also found that high-stakes grading in PI classes with clickers led to more one sided conversations as higher achieving students took over the discussion.

PI has become a widely used learning strategy. According to Fagen, Crouch, and Mazur (2002), PI has been implemented in hundreds of universities, 4-year colleges, 2-year colleges, community colleges, and high schools. In a survey of 384 teachers trying PI, more than 90% of respondents said they planned to continue using PI within their classroom (Fagen, Crouch, & Mazur, 2002).

Review of Peer Instruction.

Studies have shown statistically significant learning gains from using PI versus a traditional lecture in multiple settings. Crouch and Mazur (2001) reported the increase in student understanding of class material based on improvements on the Force Concept Inventory (FCI), the Mechanics Baseline Test, examination problems, and concept tests over a ten-year period of using PI at Harvard. Lasry, Mazur, and Watkins (2008) reported an increase in learning gains on FCI scores in their study at a two-year college similar to what Mazur reported at Harvard. The study also reported that the PI sections had fewer students drop the course than other sections taught with traditional methods. Cummings and Roberts (2008) compared learning gains of high school physics students in PI classes to those gains in traditional classes. PI lead to a 40% normalized gain in conceptual understanding while the control group only showed a 24% normalized gain.

Nicol and Boyle (2003) found that students preferred PI over class-wide discussions in a large university setting based on student interviews and responses to surveys. Smith, Wood, Krauter, and Knight (2011) found that concept questions with peer-led discussion and a teacher explanation outperformed either peer-led discussion or teacher explanation alone in large university genetics classes. Kalman, Miner-Bolotin, and Antimirova (2010) did find that collaborative group projects were more effective than PI, but the study still recommended the use of PI when transferring away from a lecture style class. The study also reported that the collaborative group project took significantly more class time, which meant that it was only used to teach two topics during the semester. The rest of the topics were still covered with PI. PI provides a good opportunity to expand newly learned concepts from the lab into more applications with less class time.

Hypotheses

The hypothesis for the first two research questions was that both peer-led and teacher-led discussions during PI would cause a statistically significant learning gain because both treatments were being tested against physics students' prior knowledge, which has been shown to contain non-Newtonian concepts (Halloun & Hestenes, 1985). So, an effective treatment should have produced a statistically significant learning gain. The alternative hypothesis was that one or both of the treatments did not show a statistically significant learning gain, which would be supported by the idea that it is difficult to overcome preconceptions (Brumby, 1984; Mazur, 1997; McDermott, 1984; West & Pines, 1985).

The hypothesis for the third research question (a comparison of peer-led and teacher-led discussions) for this study was that students who are part of the peer-led discussion approach would show greater learning gains than the teacher-led class discussion group as evidenced by more improvement on the pretest and posttest scores. According to constructivism, concepts are constructed by the learner, which results in a deeper level of understanding. Given the right experiences that cause disequilibrium, the students should be able to accommodate new mental structures for developmentally appropriate material (Bybee & Sund, 1990).

One alternative hypothesis of the third research question was that students who were part of the teacher-led class discussions would show greater learning gains than the student-led discussions as evidenced by more improvement on the pretest and posttest scores. This result could also be supported by constructivism. Vygotsky's social constructivism model is driven by the idea that interactions drive knowledge construction (Powell & Cody, 2009). Another alternative hypothesis of the third research question was that there will be no statistically significant difference measured between the peer-led and teacher-led discussions. This outcome could show that both strategies were equally effective in improving the learning gains of the students. A wide range of strategies can be effective in helping students. Marzano (2009) stated, "The entire constellation of strategies is necessary for a complete view of effective teaching" (p. 31). However, the lack of a statistically significant difference could also be caused by limitations in the study itself, like the limited research time to influence student understanding.

Chapter III-Methods

This quasi-experimental study examined the learning gains of high school physics students comparing their involvement in differing discussion approaches within Peer Instruction during the Fall 2011 school semester. For this study, these varying approaches included peer-led discussions and teacher-led discussions. For the full research timeline of activities involved in the study refer to Appendix A.

The sample for this study was students enrolled in high school physics in a large, suburban high school in the South Central region of the U.S. The total enrollment in grades nine through twelve at this school was approximately 2,100 students with an average teacher to student ratio of 1:18. Fifty-one percent of the students in this school were male while 49% were female. The student population was 79% White, 7% American Indian, 6% Hispanic, 5% Black, and 3% Asian. Twenty-six percent of the total school population was eligible for free or reduced lunch. The daily class schedule for the school was divided into seven, 55-minute periods with three morning classes before lunch and three afternoon classes after lunch.

Recruitment Methods

All students who were recruited for this research were enrolled in one of four physics classes taught by the principal investigator during the 2011-2012 school year. Two classes were taught in the morning before lunch while the other two classes were taught in the afternoon after lunch. Students in the two morning classes were defined as Group 1 and participated in peer-led discussions; students in the afternoon classes were defined as Group 2 and participated in teacher-led discussions (these distinct discussion models will be discussed later). Peer-led discussions (Group 1) were purposefully selected to take place in the morning classes to help the principal investigator/classroom teacher determine how much time would be devoted to each concept question and to identify what problems the students were having with the questions to improve the teacher-led discussions in the afternoon classes (Group 2). This strategy ensured that Group 1 and Group 2 had equal amounts of time for the concept questions. On days when peer-led and teacher-led discussions did not take place, all classes in both groups covered the same curriculum in the same manner. All students were given equal time and the same presentation of all labs, readings, worksheets with problems, and other discussions.

Ninety-four students grades ten through twelve were recruited in the study. Fifty-five students were recruited for Group 1 and 39 students were recruited for Group 2. Parental consent and student assent forms (see Appendix B) were sent home with the students on the first day of the Fall 2011 school semester. The study was introduced to the students by a neutral party, who handed out and collected the parental consent forms and the student assent forms while the principal investigator was not in the room. The forms clearly stated that participation was voluntary and students (or parents) who did not agree to participate would experience the exact same curriculum and testing. The consent and assent forms only gave the researcher permission to use the data that were collected for the study along with information about the participants' gender, GPA, and mathematics courses taken in school.

The principal investigator, as the classroom teacher, had access to all of the student information previously listed. For the purpose of the study, only the students' gender, GPA, mathematics courses taken, and data collected from the study were analyzed. All of the student names were removed from the data once the pretest and posttest data were paired together for each student. During the data collection process, the data were stored in the district's electronic grade book, which was password protected and on a thumb drive that was kept in a locked file cabinet. To address researcher bias, the principal investigator did not have access to the list of students who were participating in the study until after all of the data had been collected at the end of the study.

Seventy-one students provided parental consent/student assent and participated in the research study. There were 43 participants in Group 1 and 28 participants in Group 2. Table 1 below summarizes the participant demographics. Fifty-eight percent of the participants in Group 1 were male, and 42% were female. Fifty-seven percent of the participants in Group 2 were male, and 43% were female. For GPA students were classified into five categories. Category 1 GPA ranged from 4.0-3.75, Category 2 ranged from 3.74-3.5, Category 3 ranged from 3.49-3.25, Category 4 ranged from 3.24-3.0, and Category 5 GPA was less than 3.0. The GPA classification was chosen because the majority of the physics students had high GPAs. Sixty-five percent of Group 1 was in the highest GPA category, while only 7% of Group 1 had GPAs below 3.0. Forty-six percent of Group 2 was in the highest GPA category, while only 18% of Group 2 had GPAs below 3.0. The Current Math Class represented the current math class of the student rather than the total number of math classes taken. Students placed in Category 1 were currently in Algebra II, Category 2 students were currently in Math Analysis, Category 3 students were currently in Calculus, and Category 4 students had already completed Calculus. Forty-two percent of participants in Group 1 had not taken

Calculus, while 58% of the participants in Group 1 were at least currently enrolled in Calculus. Forty-six percent of participants in Group 2 had not taken Calculus, while 54% of the participants in Group 2 were at least currently enrolled in Calculus.

Table 1

Participant Demographics

		Gender		GPA					Current Math Class			
Group	*n*	Male	Female	1	2	3	4	5	1	2	3	4
1	43	25	18	28	4	5	3	3	2	16	19	6
2	28	16	12	13	3	4	3	5	2	11	13	2
Total	71	41	30	41	7	9	6	8	4	27	32	8

Note. GPA – 1 = 4.0-3.75, 2 = 3.74-3.5, 3 = 3.49-3.25, 4 = 3.24-3.0, 5 = < 3.0, Current Math Class – 1 = Algebra II, 2 = Math Analysis, 3 = Calculus, 4 = Beyond Calculus

Instruments Used

This study used the Force Concept Inventory (FCI) (Hestenes, Wells, & Swackhamer, 1992) and physics concept questions and benchmark tests from the Peer Instruction (PI) model (Mazur, 1997) to measure students' learning gains throughout the duration of the study. Deeper understanding of the physics concepts was indicated by higher scores on the FCI, concept questions, and benchmark tests. An informal journal was also kept by the principal investigator to take notes on the questions covered, duration of each question, and any other applicable information during the days when data were collected in the classroom. This information was used to provide general observations for the discussion section of the thesis.

Force Concept Inventory (FCI).

The Force Concept Inventory (FCI) (Hestenes, Wells, & Swackhamer, 1992) is available as a .pdf file and may be accessed online by requesting a password (Koch, 2009). Due to accessibility restrictions, the authors of the FCI prefer that the instrument remains secure; therefore, the FCI is not provided in this document. The FCI consists of 30 multiple choice questions, and each question presents students with an option consistent with the Newtonian understanding of the concept and four alternative options that are consistent with common misconceptions that are based on commonsense explanations of experiences. The FCI was designed to measure conceptual, not computational, understandings of Newtonian mechanics.

Hestenes and Halloun (1995) stated that scoring below 60% on the FCI represents three flaws in students' understanding: (a) the individual does not make a clear distinction between variables, like velocity and acceleration; (b) the individual does not realize that changes in the motion of an object are only caused by forces; and (c) the individual does not have a consistent system of understanding motion and forces. Scoring between 60% and 85% represents the beginnings of a consistent system of understanding the concept of force, and scoring above 85% represents a completely consistent system agreeing with the Newtonian concept of force. Hestenes, Wells, and Swackhamer (1995) summarized the concepts covered by each of the items on the FCI (see Table 2).

Table 2

Newtonian Concepts on the Force Concept Inventory

	Inventory Item
0. Kinematics	
Velocity discriminated from position	20E
Acceleration discriminated from velocity	21D
Constant acceleration entails	
parabolic orbit	23D, 24E
changing speed	25B
Vector addition of velocities	(7E)
I. First Law	
with no force	4B, (6B), 10B
velocity direction constant	26B
speed constant	8A, 27 A
with cancelling forces	18B,28C
2. Second Law	
Impulsive force	(6B), (7E)
Constant force implies constant acceleration	24E, 25B
3. Third Law	
for impulsive forces	2E, llE
for continuous forces	13A, 14A
4. Superposition Principle	
Vector sum	19B
Cancelling forces	(9D), 18B, 28C
5. Kinds or Force	
5S. Solid contact	
passive	(9D), (12 B,D)
Impulsive	15C
Friction opposes motion	29C
5F. Fluid contact	
Air resistance	22D
buoyant (air pressure)	12D
5G. Gravitation	5D, 9D, (12B,D), 17C, I8B, 22D
acceleration independent of weight	1C, 3A
parabolic trajectory	16B, 23D

Note. The image of the list of concepts included on the Force Concept Inventory is from http://modeling.asu.edu/r%26e/fci.pdf.

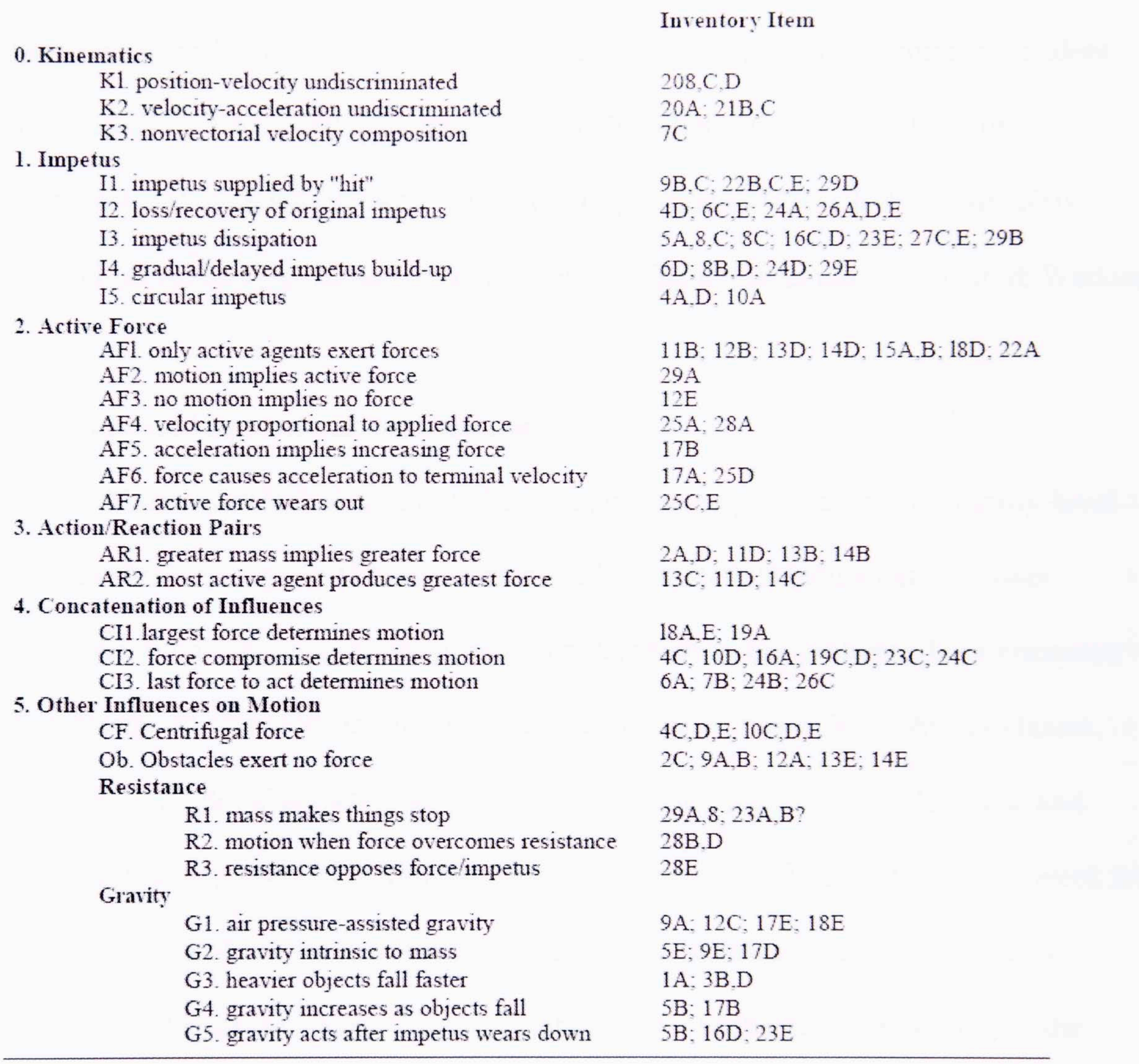
Over half of the questions on the FCI were originally included in the Mechanics Diagnostic Test (MDT) (Halloun & Hestenes, 1985). Halloun and Hestenes showed that the MDT had a Kuder-Richardson reliability coefficient of 0.86 for pre-test scores and 0.89 for post-test scores showing the high reliability of the test. The reliability and test validity of the FCI has not yet been determined (Savinainen & Scott, 2002).

Even though formal tests of reliability and validity have not been conducted for the FCI, Hestenes, Halloun, Heller, and Huffman reported that the FCI is an effective test to assess learning strategies and to evaluate a class' understanding (as cited in Saul, 1998). Hake (1998) reported the consistency between FCI and MDT pre-test and post-test scores for over 6,000 high school and university students from over sixty different introductory physics classes. It was determined that similar classes will produce similar pre-test and post-test scores on the FCI (Saul, 1998).

The controversy over the FCI initiated by Huffman and Heller (1995) questioned what the FCI actually measured. Analyzing FCI responses from 750 university and 145 high school students, the researchers found no correlation between scores from questions over the same concept. Huffman and Heller further questioned whether the concept of "Force" within the FCI could be broken down into the six constituent concepts. They proposed that the test measured isolated ideas, which then did not evaluate the students overall understanding of the concept. Halloun and Hestenes (1995) responded that the six concepts of force on the FCI were meant to be a standard to evaluate student understanding of the concept of force. They emphasized the need to look at the results from the entire test to evaluate student understanding of Newtonian concepts. The six concepts of force on the FCI were not meant to represent how students built knowledge about the concept of force, but merely to evaluate the students' understanding of different aspects of the concept to help identify different non-Newtonian concepts held by the class. Hestenes, Wells, and Swackhamer (1995) summarized the connection between common misconceptions as revealed by the six force-related concepts and questions on the FCI (see Table 3). Presence of the misconceptions was suggested by selection of the corresponding FCI item.

Table 3

A Taxonomy of Misconceptions by the Force Concept Inventory



	Inventory Item
0. Kinematics	
K1. position-velocity undiscriminated	208,C,D
K2. velocity-acceleration undiscriminated	20A; 21B,C
K3. nonvectorial velocity composition	7C
1. Impetus	
I1. impetus supplied by "hit"	9B,C; 22B,C,E; 29D
I2. loss/recovery of original impetus	4D; 6C,E; 24A; 26A,D,E
I3. impetus dissipation	5A,8,C; 8C; 16C,D; 23E; 27C,E; 29B
I4. gradual/delayed impetus build-up	6D; 8B,D; 24D; 29E
I5. circular impetus	4A,D; 10A
2. Active Force	
AF1. only active agents exert forces	11B; 12B; 13D; 14D; 15A,B; 18D; 22A
AF2. motion implies active force	29A
AF3. no motion implies no force	12E
AF4. velocity proportional to applied force	25A; 28A
AF5. acceleration implies increasing force	17B
AF6. force causes acceleration to terminal velocity	17A; 25D
AF7. active force wears out	25C,E
3. Action/Reaction Pairs	
AR1. greater mass implies greater force	2A,D; 11D; 13B; 14B
AR2. most active agent produces greatest force	13C; 11D; 14C
4. Concatenation of Influences	
CI1.largest force determines motion	18A,E; 19A
CI2. force compromise determines motion	4C, 10D; 16A; 19C,D; 23C; 24C
CI3. last force to act determines motion	6A; 7B; 24B; 26C
5. Other Influences on Motion	
CF. Centrifugal force	4C,D,E; 10C,D,E
Ob. Obstacles exert no force	2C; 9A,B; 12A; 13E; 14E
Resistance	
R1. mass makes things stop	29A,8; 23A,B?
R2. motion when force overcomes resistance	28B,D
R3. resistance opposes force/impetus	28E
Gravity	
G1. air pressure-assisted gravity	9A; 12C; 17E; 18E
G2. gravity intrinsic to mass	5E; 9E; 17D
G3. heavier objects fall faster	1A; 3B,D
G4. gravity increases as objects fall	5B; 17B
G5. gravity acts after impetus wears down	5B; 16D; 23E

Note. The image of the taxonomy of misconceptions probed by the Force Concept Inventory is from http://modeling.asu.edu/r%26e/fci.pdf.

Concept questions and benchmark tests.

The concept questions and the benchmark test items for this study were both taken from *Peer Instruction: A User's Manual* (Mazur, 1997). A full list of the questions and explanations can be found in the book. Each set of concept questions covered one of the following physics concepts: one-dimensional motion, free fall, horizontal projectile motion, projectile motion with an angle, Newton's Laws in One-Dimension (1-D), and Newton's Laws in Two-Dimensions (2-D). The concept questions served as a measure to gauge student progress and to help improve student understanding. No formal tests were done on the reliability and validity of the questions, but the questions were used in multiple studies (Crouch & Mazur, 2001; Cummings & Roberts, 2008; Fagen, Crouch, & Mazur, 2002; Lasry, Mazur, & Watkins, 2008).

Peer Instruction Intervention Rationale

Mazur (1997) designed PI to be included in every lecture in a university-level physics class. He required his students to read the materials before class, answer questions online about the reading, and complete homework problems for subsequent discussions. PI was modified for this study with regular high school physics classes, which had smaller class sizes, more instructional time, more concrete learners, and different learning objectives. Students meet in high school classes five times a week for an entire year rather than just three lectures and a discussion class per week for a semester in a typical university class. The high school schedule allows more of the material to be covered multiple times in class with different strategies to help more concrete learners. For this study, the pre-class reading and online questions as typified by Mazur at the university level were replaced with inquiry-based labs, discussions, and conceptual and mathematical problems.

Peer Instruction Approaches

Concept questions.

For this study, the concept questions were grouped together after the concepts were experienced in the lab. Each concept question discussion session began with a reminder of the importance of participating in the discussions to develop a systematic understanding of the physics principles rather than just a disconnected list of physics facts. Both groups were given the same set of concept questions, and if time was a limitation, the questions were either posted online or finished the following class period.

Immediate feedback.

Each student was given approximately one minute to read the question and enter a response using an individual clicker within a classroom set of clickers. The immediate feedback was used to gauge student understanding of the concept questions and to guide the teacher as to what decisions should be made subsequently. If more than 75% of the students answered a question correctly, the teacher revealed the results and asked the students to explain why that choice was consistent with the Newtonian concept. If fewer than 35% of the students answered a question correctly, another demonstration or lecture was given before asking the original question again (Fagen, Crouch, & Mazur, 2002). If between 35% and 75% of the class answered a question correctly, then the question was used for one of two discussion formats (peer-led or teacher-led). This study focused on whether using peer-led or teacher-led discussions at this point was more effective at increasing learning gains. Figure 1 summarizes the steps of each of the discussion formats that were used in the study that will be discussed in the next section.

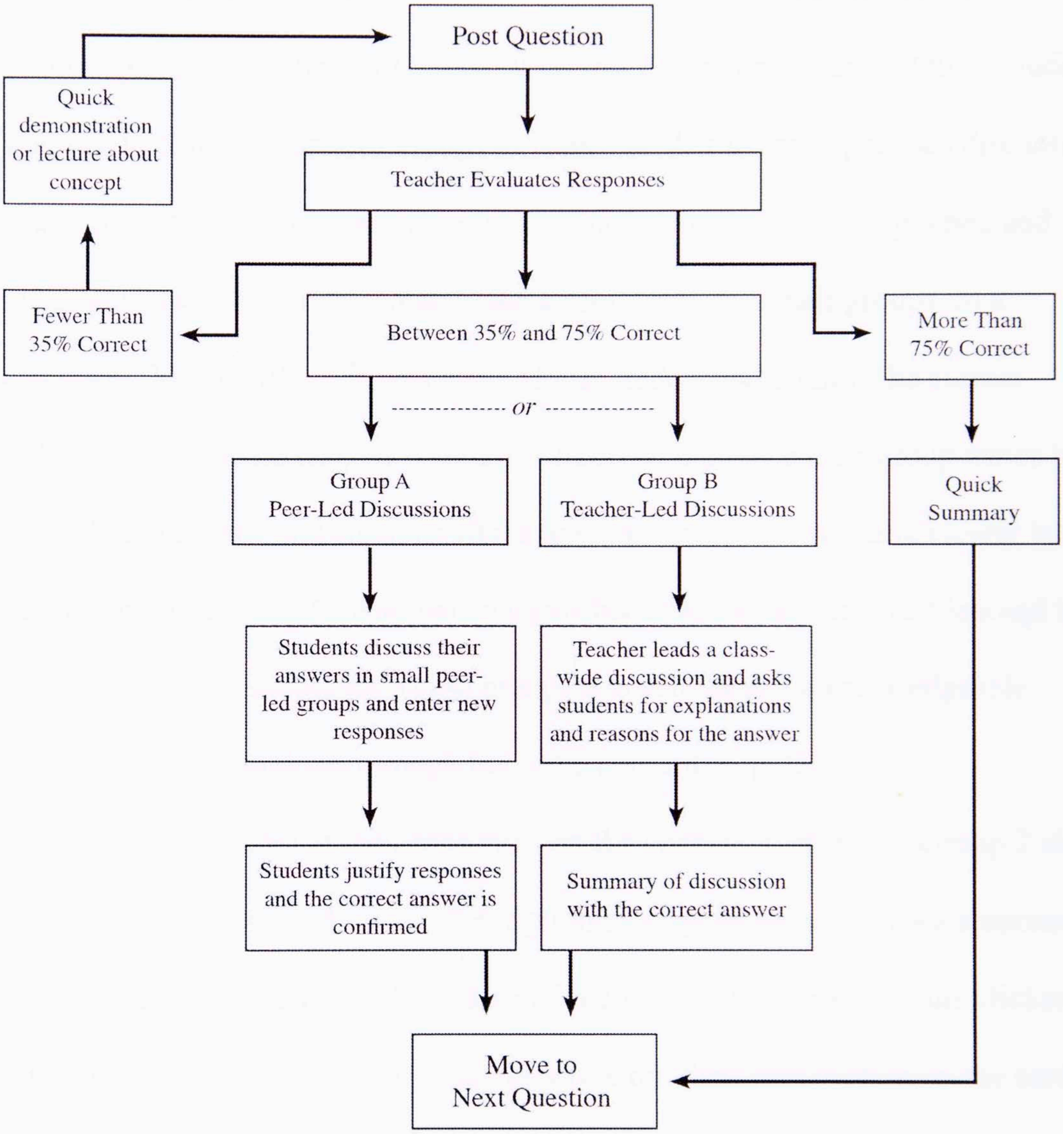


Figure 1. Sequence of peer-led and teacher-led discussions used in the study.

Peer-led or teacher-led discussions.

When 35% to 75% of the students in Group 1 had individually answered a concept question correctly using the clicker system, they followed the peer-led discussion sequence described by Crouch, Watkins, Fagen, and Mazur (2007). Students were allowed to choose their own groups to work together in small peer-led discussions for three to four minutes to answer the question again. Students then explained and justified their answers for approximately three minutes within their groups. In a previous study (Mazur, 1997), it was realized that students who knew the correct answer for the appropriate reasons were more likely to convince their group mates to reconsider their answers, and students who knew the correct answer have clearer insight into their group mates' non-Newtonian concepts because they all have just learned to work through the same problems. These groups allowed the more knowledgeable students to walk other students through the same conceptual pitfalls.

In addition to entering their responses on the clickers, students in Group 2 also wrote their answer choice and a short description or diagram to justify their response on a slip of paper anonymously. Students entered their initial response with the clickers during the first minute, and then the students wrote out their justification on the card during the time when Group 1 would have broken into small group discussions. When 35% to 75% of the students individually answered a concept question correctly, Group 2 used the following steps for teacher-led discussions. The teacher quickly gathered the slips of paper and wrote a few of the different responses on the board, placed the slips of paper under the document camera, or read them aloud. Students were given two minutes to think about why each of the other choices contained non-Newtonian concepts. The teacher then asked for alternative choices or justifications from the class. For two minutes, all justifications were organized on the whiteboard under the corresponding response choice letters. Once all choices and justifications were on the board, a second row was added to the board. The teacher asked the students for non-Newtonian ideas in any of the justifications to be written on the second row. Once all of the non-Newtonian ideas were organized on the whiteboard, the teacher asked the students to vote on an option by a show of hands. Whichever group had the largest number of votes responded to challenges of non-Newtonian ideas in their column and explained for three to four minutes why each of the other options did not agree with the Newtonian concept. If the majority of students did not have the correct Newtonian explanation, the teacher explained why that choice was non-Newtonian, and the class voted again on the remaining options.

The procedure for Group 2 was being proposed by the principal investigator to test an alternative to peer-led discussions within PI. Group 2 still followed many of the same effective steps from the typical PI method. Group 2 answered the same concept questions with the clickers, and the teacher still used the results to determine whether the class would receive another lecture or demonstration, to quickly summarize the answer and move on to the next question, or to have a discussion about the question. Having the students write down a justification for their response and think about why the other choices were non-Newtonian helped the students reflect on the concepts. Encouraging the students to reflect and ask questions helped the students develop a better understanding of the concepts (Watts & de Jesus, 2005). According to Watts and de Jesus (2005), "questioning prompted by a questioner's own reflections leads to deeper, more fundamental questions than that prompted by external factors such as peers, textbooks, or teachers" (p. 439). Studies have also shown that self-reflective questions are able to have a lasting effect on student conceptions (King, 1992; Rosenshine, Meister, & Chapman, 1996). So, having students reflect on the choices may be another effective type of discussion other than peer-led discussions.

Procedures

This study analyzed the effectiveness of two different discussion techniques for incorporating Peer Instruction in a high school physics course. Effectiveness was measured by learning gains on pretest and posttest FCI scores as well as between sets of concept questions and corresponding benchmark tests. See Appendix A for an itemized research timeline.

The pretest and posttest of the FCI was given to both groups in the same way. A hard copy of the FCI test was given to each student, and the students entered their responses with the clickers at their own pace. The pretest was administered at the beginning of the semester and the posttest was administered eleven weeks later to determine students' learning gains of conceptual understanding of Newtonian mechanics. The FCI required one fifty minute class period each time it was administered.

The concept questions were administered six times throughout the duration of the research study. The time between administrations varied between one and two weeks. The concept questions served as a measure to gauge student progress and to help improve student understanding through either peer-led (Group 1) or teacher-led (Group 2) discussions. Answering and reviewing each set of concept questions took approximately thirty minutes of class time with both groups. Both groups had a low-stakes grading system for the sets of concept questions. Willoughby's (2009) study of grading incentive with clickers showed that high-stakes grading discouraged group discussions as the stronger academic students became more dominant, which allowed the lower achieving students to passively block vote with the group. For this study, students in Group 1 and Group 2 were given a participation grade for individually answering the questions with the clickers before the peer-led or teacher-led discussions.

After students completed two sets of concept questions, they were administered a benchmark test that measured their conceptual understandings of the inherent physics concepts within the preceding set of concept questions. Throughout the study, students completed three benchmark tests. The time between administrations varied between two and three weeks. The benchmark scores were compared to the concept questions scores to determine learning gains over the inherent physics concepts. Each benchmark test took approximately twenty minutes of class time and was administered to Group 1 and Group 2 in the same way. Students were given a hard copy of benchmark questions to work through at their own pace, and the students entered their responses on the clickers. One participation grade was given for each two sets of concept questions for both groups.

Chapter IV-Results

Summary of Assessment Scores

Table 4 lists the number of participants who completed each assessment along with the number of questions on each assessment and the mean percentage for each assessment for each group. The reason why the number of participants varies is because some students may have been absent or pulled out of class for an activity on different days when the discussions took place. A participant's data were not included unless they were present for the pretest and posttest assessments. The same numbers of questions were asked to each group, but the number of questions covered in each section was based on the concept covered and the amount of class time available, which means that each assessment had a different number of questions. Each set of Concept Questions (CQ) had between two and four questions and served as a pretest. Each Benchmark (B) covered two sets of concept questions, and so is referred to as either A or B for the first or second set of concepts. The benchmarks ranged between three and five questions for each of the topics. The FCI had 30 questions, and it was administered as a pretest and posttest. The mean scores in Table 4 are reported as percentages for each of the assessments, and the standard deviation is reported along with each mean percentage. The mean percentages for Group 1 on the pretest scores ranged from 12% to 41%, and the mean percentages for Group 2 on the pretest scores ranged from 26% to 44%. The mean percentages for Group 1 on the posttest scores ranged from 45% to 78%, and the mean percentages for Group 2 on the posttest scores ranged from 52% to 82%.

Table 4

Summary of Assessment Scores

			Pretest Questions			Posttest Questions		
Group	Assessment	*n*	#Q	*M*	*SD*	#Q	*M*	*SD*
1	CQ1-B1A	42	4	32.1	23.6	5	56.7	26.8
	CQ2-B1B	43	3	35.6	22.9	5	74.9	16.4
	CQ3-B2A	34	2	41.1	19.3	3	73.5	25.6
	CQ4-B2B	43	3	11.6	19.0	3	78.3	26.1
	CQ5-B3A	43	3	40.3	22.4	4	61.6	26.9
	CQ6-B3B	43	3	37.2	23.2	4	51.7	18.4
	FCI	43	30	23.6	8.8	30	45.0	16.0
2	CQ1-B1A	27	4	38.9	26.3	5	68.9	23.8
	CQ2-B1B	24	3	44.4	23.4	5	76.7	19.3
	CQ3-B2A	20	2	40.4	20.5	3	81.7	27.5
	CQ4-B2B	28	3	27.4	25.7	3	82.1	23.1
	CQ5-B3A	28	3	45.2	20.7	4	58.9	19.5
	CQ6-B3B	27	3	35.8	22.5	4	57.4	20.6
	FCI	28	30	25.9	11.9	30	51.8	14.8

Note. CQ = Concept Questions; B = Benchmark test; FCI = Force Concept Inventory; #Q = Number of questions on each assessment; mean = Mean percentage of each assessment.

Grouping of FCI Scores

Table 5 summarizes the number of students that scored low, moderate, and high on the FCI pretest and posttest based on Hestenes and Halloun's (1995) analysis of FCI scores. Students who answered fewer than 18 out of the 30 questions correctly had a *low* understanding of the distinction between variables, the concept that forces cause motion, and how to apply the principles consistently. Students who answered between 18 and 24 of the questions correctly had a *moderate* understanding, which meant that they had the beginnings of a consistent system of understanding the concepts of forces and motion. Students who answered more than 24 questions correctly were in the *high* category, which meant that they had a completely consistent system of the Newtonian concept of force. In this study, all participants scored low on the pretest FCI. Ten students scored moderate in both Group 1 and Group 2 on the posttest FCI, but no one scored high in either group on either assessment.

Table 5

Grouping of FCI Scores

			Number of Students in Each Range of Scores		
Group	FCI Score	*n*	Low $S < 18$	Moderate $18 < S < 24$	High $S > 24$
1	Pretest	43	43	0	0
	Posttest	43	33	10	0
2	Pretest	28	28	0	0
	Posttest	28	18	10	0

Note. S = Score on FCI out of 30.

Paired Samples T-test

To answer research questions one and two, a paired samples t-test was used to compare learning gains in peer-led discussions (Group 1) and teacher-led discussions (Group 2). This procedure was done with the pretest and posttest for the FCI for the entire study and for the pretest concept questions and posttest benchmark questions for each of the concept sections. The paired samples t-test for Group 1 showed if there was a statistically significant increase between the pretest and posttest percentage scores when using concept questions (CQ) with peer-led discussions. The paired samples t-test for Group 2 showed if there was a statistically significant increase between the pretest and posttest scores when using concept questions with teacher-led discussions.

The paired samples t-test was conducted for Group 1 to compare the students' conceptual understanding of forces at the time of the pretest and the time of the posttest. All seven pairs of assessments for Group 1 showed a statistically significant increase between the pretest and posttest scores. There was a significant difference in the scores for Concept Question 1 ($M = 32.1$, $SD = 23.6$) and Benchmark 1A ($M = 56.7$, $SD = 26.8$); $t(41) = 5.263$, $p = .000$. There was a significant difference in the scores for Concept Question 2 ($M = 35.6$, $SD = 22.9$) and Benchmark 1B ($M = 74.9$, $SD = 16.4$); $t(42) = 9.881$, $p = .000$. There was a significant difference in the scores for Concept Question 3 ($M = 41.1$, $SD = 19.3$) and Benchmark 2A ($M = 73.5$, $SD = 25.6$); $t(33) = 6.640$, $p = .000$. There was a significant difference in the scores for Concept Question 4 ($M = 11.6$, $SD = 19.0$) and Benchmark 2B ($M = 78.3$, $SD = 26.1$); $t(42) = 17.349$, $p = .000$. There was a significant difference in the scores for Concept Question 5 ($M = 40.3$, $SD = 22.4$) and Benchmark 3A ($M = 61.6$, $SD = 26.9$); $t(42) = 4.107$, $p = .000$. There was a significant difference in the scores for Concept Question 6 (*M* = 37.2, *SD* = 23.2) and Benchmark 3B (*M* = 51.7, *SD* = 18.4); *t*(42) = 3.341, *p* = .002. There was a significant difference in the scores for the FCI Pretest (*M* = 23.6, *SD* = 8.8) and the FCI Posttest (*M* = 45.0, *SD* = 16.0); *t*(42) = 9.989, *p* = .000. The information is summarized in Table 4 and Table 6. The change between the pretest and posttest score is referred to as the difference score in Table 6.

The paired samples t-test was conducted for Group 2 to compare the students' conceptual understanding of forces at the time of the pretest and the time of the posttest. All seven pairs of assessments for Group 2 show a statistically significant increase between the pretest and posttest scores. There was a significant difference in the scores for Concept Question 1 (*M* = 38.9, *SD* = 26.3) and Benchmark 1A (*M* = 68.9, *SD* = 23.8); *t*(26) = 5.002, *p* = .000. There was a significant difference in the scores for Concept Question 2 (*M* = 44.4, *SD* = 23.4) and Benchmark 1B (*M* = 76.7, *SD* = 19.3); *t*(23) = 6.111, *p* = .000. There was a significant difference in the scores for Concept Question 3 (*M* = 40.4, *SD* = 20.5) and Benchmark 2A (*M* = 81.7, *SD* = 27.5); *t*(19) = 5.349, *p* = .000. There was a significant difference in the scores for Concept Question 4 (*M* = 27.4, *SD* = 25.7) and Benchmark 2B (*M* = 82.1, *SD* = 23.1); *t*(27) = 10.522, *p* = .000. There was a significant difference in the scores for Concept Question 5 (*M* = 45.2, *SD* = 20.7) and Benchmark 3A (*M* = 58.9, *SD* = 19.5); *t*(27) = 2.821, *p* = .009. There was a significant difference in the scores for Concept Question 6 (*M* = 35.8, *SD* = 22.5) and Benchmark 3B (*M* = 57.4, *SD* = 20.6); *t*(26) = 4.030, *p* = .000. There was a significant difference in the scores for the FCI Pretest (*M* = 25.9, *SD* = 11.9) and the FCI Posttest (*M* = 51.8, *SD* = 14.8); *t*(27) = 9.556, *p* = .000. The information is summarized in Table 4 and Table 6. The change between the pretest and posttest score is referred to as the difference score in Table 6.

Table 6

Significance of the Difference Scores for Each Pair of Assessments Using the Paired Samples T-test

Group	Assessment	*M*	*SD*	*t*	df	*p*
1	CQ1-B1A	24.5	30.2	5.263	41	.000
	CQ2-B1B	39.2	26.0	9.881	42	.000
	CQ3-B2A	33.3	28.4	6.640	33	.000
	CQ4-B2B	66.7	25.2	17.349	42	.000
	CQ5-B3A	21.3	34.0	4.107	42	.000
	CQ6-B3B	14.5	28.5	3.341	42	.002
	FCI_{post}-FCI_{pre}	21.3	14.0	9.989	42	.000
2	CQ1-B1A	30.0	31.1	5.002	26	.000
	CQ2-B1B	32.2	25.8	6.111	23	.000
	CQ3-B2A	41.7	34.8	5.349	19	.000
	CQ4-B2B	54.8	27.5	10.522	27	.000
	CQ5-B3A	13.7	25.7	2.821	27	.009
	CQ6-B3B	21.6	27.9	4.030	26	.000
	FCI_{post}-FCI_{pre}	25.8	14.3	9.556	27	.000

Note. Mean = Average change between the pretest and posttest percentage scores.

Table 7 shows the correlation between the pretest scores and the posttest scores from the paired samples t-test. Most of the correlations were small. The only moderate correlations between the students taking the pretest and the students taking the posttest were the FCI pretest and posttest and Concept Question 4 to B2B. This correlation was moderate for both Group 1 and Group 2 for these pairs of assessments. Most of the correlations were not significant. There were only three statistically significant correlations. Concept Question 4 to Benchmark 2B for Group 1 had a *p*-value of .006. The FCI pretest and posttest for Group 1 had a *p*-value of .001. The FCI pretest and posttest for Group 2 had a *p*-value of .018. In addition, Concept Question 4 to Benchmark 2B for Group 2 was approaching significance with a *p*-value of .054.

Table 7

Correlation of Each Pair of Assessments Using the Paired Samples T-test

Group	Assessment	*n*	Correlation	Size	*p*
1	CQ1-B1A	42	.289	Small	.064
	CQ2-B1B	43	.120	Small	.442
	CQ3-B2A	34	.227	Small	.196
	CQ4-B2B	43	.412	Moderate	.006
	CQ5-B3A	43	.059	Small	.705
	CQ6-B3B	43	.077	Small	.626
	FCI_{post}-FCI_{pre}	43	.491	Moderate	.001
2	CQ1-B1A	27	.226	Small	.257
	CQ2-B1B	24	.279	Small	.187
	CQ3-B2A	20	-.031	Small	.897
	CQ4-B2B	28	.368	Moderate	.054
	CQ5-B3A	28	.186	Small	.345
	CQ6-B3B	27	.167	Small	.406
	FCI_{post}-FCI_{pre}	28	.444	Moderate	.018

Note. CQ = Concept Question; B = Benchmark; Size = Strength of the Correlation.

Relationship of Gender, GPA and Math Level to FCI Pretest

A one-way between-subjects ANOVA was conducted to compare the effect of gender on FCI pretest scores. The within-subjects effect referred to the change in FCI pretest scores within each separate gender, and the between-subjects effect referred to the difference between the average male and average female FCI pretest score. Gender had no significant effect on FCI pretest scores for Group 1 as measured between males (*M* = 7.48, *SD* = 3.03) and females (*M* = 6.56, *SD* = 1.98); *F*(1, 41) = 1.28, *p* = .265. Gender had a significant effect on FCI pretest scores for Group 2 as measured between males (*M* = 9.40, *SD* =.912) and females (*M* = 4.81, *SD* = 1.24); *F*(1, 26) = 7.86, *p* = .009. Males scored significantly higher on the FCI pretest than females in Group 2. The descriptive statistics of gender on FCI pretest scores for Group 1and Group 2 are in Table 8, and the ANOVA results for the effects of gender on FCI pretest scores for Group 1and Group 2 are in Table 9.

Table 8

Descriptive Statistics of Gender on FCI Pretest for Group 1and Group 2

Group	Gender	*n*	*M*	*SD*
1	Male	25	7.48	3.03
	Female	18	6.56	1.98
	Total	43	7.09	2.65
2	Male	16	9.40	.912
	Female	12	4.81	1.24
	Total	28	7.78	3.57

Table 9

One-Way ANOVA of Gender Effect on FCI Pretest Scores for Group 1and Group 2

Group	Effects	*SS*	df	*MS*	*F*	*p*
1	Between Group	8.94	1	8.94	1.28	.265
	Within Group	286.6	41	6.99		
	Total	295.6	42			
2	Between Group	80.0	1	80.0	7.87	.009
	Within Group	264.6	26	10.18		
	Total	344.7	27			

The one-way between-subjects effect ANOVA test shows that there is no significant effect of gender on the FCI pretest scores for Group 1. There is a significant effect of gender on the FCI pretest scores for Group 2, and the males in Group 2 score significantly higher than the females in Group 2.

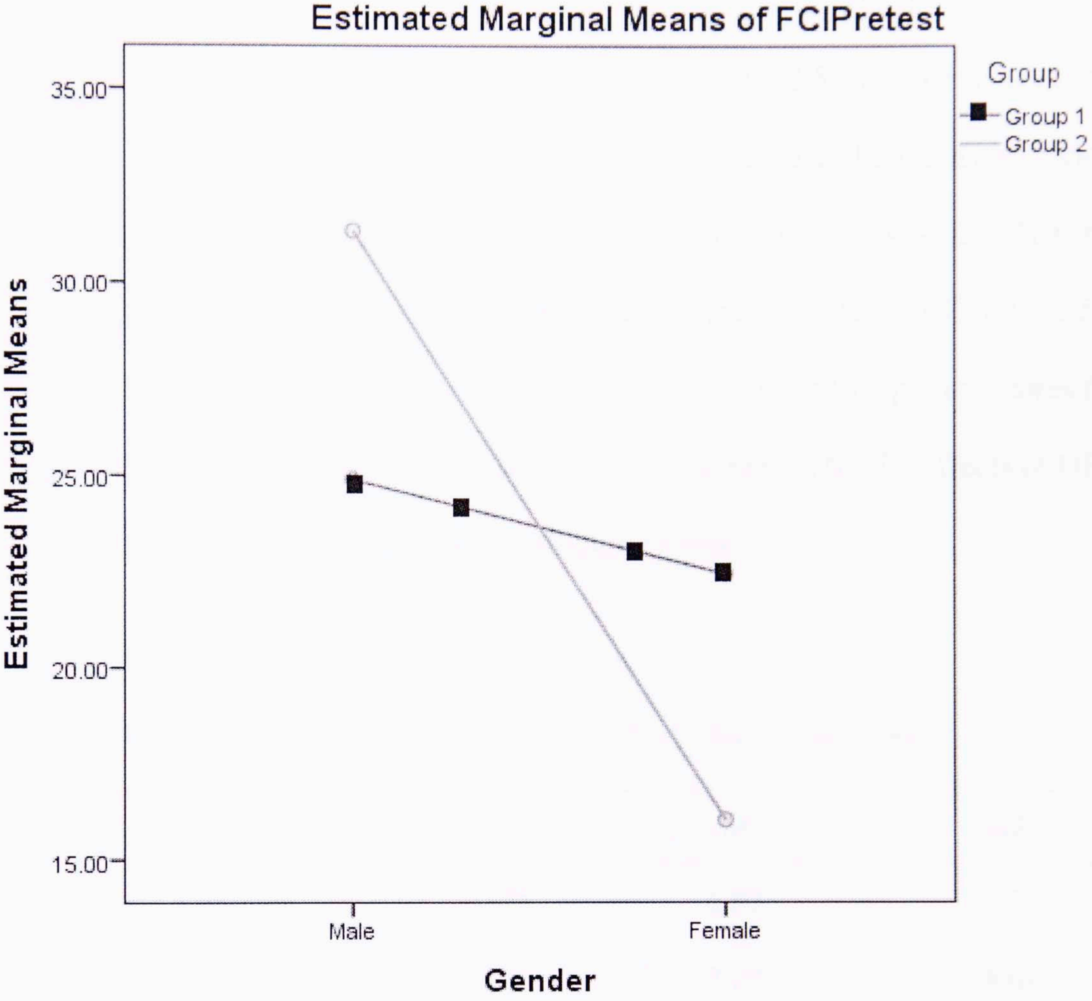


Figure 2. Plot of FCI pretest percentage scores vs. gender for Group 1 and Group 2.

A one-way between-subjects ANOVA was conducted to compare the effect of GPA on FCI pretest scores. GPA was classified into five categories. Category 1 GPA ranged from 4.0-3.75, Category 2 ranged from 3.74-3.5, Category 3 ranged from 3.49-3.25, Category 4 ranged from 3.24-3.0, and Category 5 GPA was less than 3.0. The within-subjects effect referred to the change in FCI pretest scores within each separate GPA, and the between-subjects effect referred to the differences between the categories of GPA on the FCI pretest score. GPA had no significant effect on FCI pretest scores for Group 1 as measured between categories, 4.0-3.75 ($M = 6.82$, $SD = 2.79$), 3.74-3.5

(*M* = 7.00, *SD* = .816), 3.49-3.25 (*M* = 6.00, *SD* = 1.87), 3.24-3.0 (*M* = 7.33, *SD* = 1.15), and less than 3.0 (*M* = 11.33, *SD* = 1.52); *F*(4, 38) = 2.53, *p* = .056. GPA had no significant effect on FCI pretest scores for Group 2 as measured between categories, 4.0-3.75 (*M* = 6.31, *SD* = 1.18), 3.74-3.5 (*M* = 8.67, *SD* = 1.90), 3.49-3.25 (*M* = 9.33, *SD* = 1.90), 3.24-3.0 (*M* = 8.33, *SD* = 1.90), and less than 3.0 (*M* = 7.67, *SD* = 1.50); *F*(4, 23) = .905, *p* = .477. The descriptive statistics of GPA on FCI pretest scores for Group 1and Group 2 are in Table 10, and the ANOVA results for the effects of GPA on FCI pretest scores for Group 1and Group 2 are in Table 11.

Table 10

Descriptive Statistics of GPA on FCI Pretest for Group 1and Group 2

Group	GPA	*n*	*M*	*SD*
1	4.0-3.75	28	6.82	2.79
	3.74-3.5	4	7.00	.816
	3.49-3.25	5	6.00	1.87
	3.24-3.0	3	7.33	1.15
	< 3.0	3	11.33	1.52
	Total	43	7.09	2.65
2	4.0-3.75	13	6.31	1.18
	3.74-3.5	3	8.67	1.90
	3.49-3.25	4	9.33	1.90
	3.24-3.0	3	8.33	1.90
	< 3.0	5	7.67	1.50
	Total	28	7.78	3.57

Table 11

One-Way ANOVA of GPA Effect on FCI Pretest Scores for Group 1and Group 2

Group	Effects	*SS*	df	*MS*	*F*	*p*
1	Between Group	62.2	4	15.5	2.53	.056
	Within Group	233.4	38	6.14		
	Total	295.6	42			
2	Between Group	46.9	4	11.7	.905	.477
	Within Group	297.8	23	12.9		
	Total	344.7	27			

The one-way between-subjects effect ANOVA test shows that there is no significant effect of GPA on the FCI pretest scores for Group 1 or Group 2.

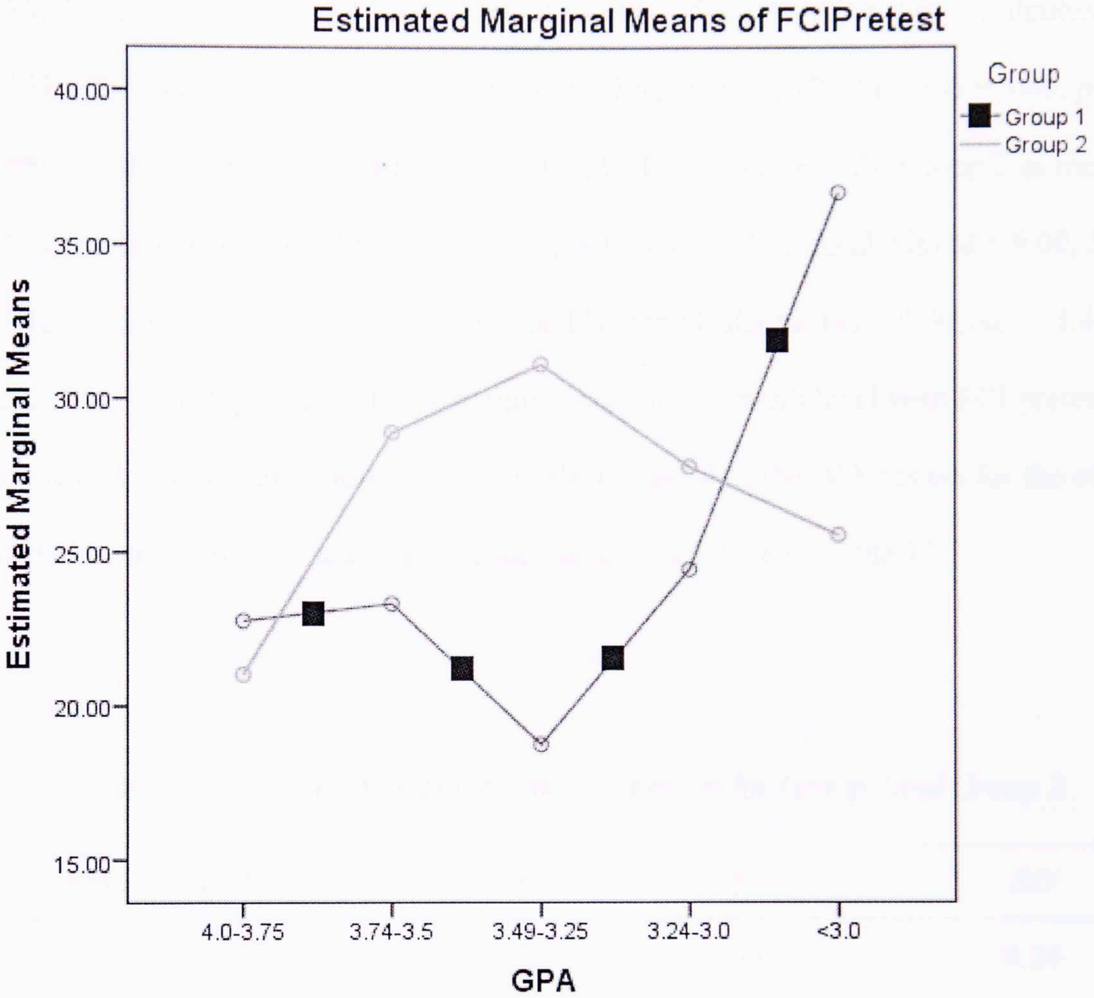


Figure 3. Plot of FCI pretest percentage scores vs. GPA for Group 1 and Group 2.

A one-way between-subjects ANOVA was conducted to compare the effect of math level on FCI pretest scores. A student's math level was classified into four categories based on their current math class. Students placed in Category 1 were currently in Algebra II, Category 2 students were currently in Math Analysis, Category 3 students were currently in Calculus, and Category 4 students had already completed Calculus. The within-subjects effect referred to the change in FCI pretest scores within each separate math level, and the between subjects effect referred to the differences between the categories of math level on the FCI pretest score. Math level had no significant effect on FCI pretest scores for Group 1 as measured between categories, Algebra II (*M* = 7.00, *SD* = 4.24), Math Analysis (*M* = 6.94, *SD* = 2.43), Calculus (*M* = 7.11, *SD* = 3.02), and Beyond Calculus (*M* = 7.50, *SD* = 2.17); *F*(3, 39) = .065, *p* = .980. Math level had no significant effect on FCI pretest scores for Group 2 as measured between categories, Algebra II (*M* = 9.00, *SD* = 4.42), Math Analysis (*M* = 9.00, *SD* = 4.12), Calculus (*M* = 7.15, *SD* = 2.91), and Beyond Calculus (*M* = 4.00, *SD* = 1.41); *F*(3, 24) = 1.454, *p* = .252. The descriptive statistics of math level with FCI pretest scores for Group 1and Group 2 are in Table 12, and the ANOVA results for the effects of GPA on FCI pretest scores for Group 1and Group 2 are in Table 13.

Table 12

Descriptive Statistics of Math Level with FCI Pretest for Group 1and Group 2

Group	Math Level	*n*	*M*	*SD*
1	Algebra II	2	7.00	4.24
	Math Analysis	16	6.94	2.43
	Calculus	19	7.11	3.02
	Beyond Calculus	6	7.50	2.17
	Total	43	7.09	2.65
2	Algebra II	2	9.00	4.42
	Math Analysis	11	9.00	4.12
	Calculus	13	7.15	2.91
	Beyond Calculus	2	4.00	1.41
	Total	28	7.78	3.57

Table 13

One-Way ANOVA of Math Level Effect on FCI Pretest Scores for Group 1and Group 2

Group	Effects	*SS*	df	*MS*	*F*	*P*
1	Between Group	1.40	3	.467	.065	.980
	Within Group	294.2	39	7.54		
	Total	295.6	42			
2	Between Group	53.0	3	17.7	1.45	.252
	Within Group	291.7	24	12.2		
	Total	344.7	27			

The one-way between-subjects effect ANOVA test shows that there is no significant effect of math level on the FCI pretest scores for Group 1 or Group 2.

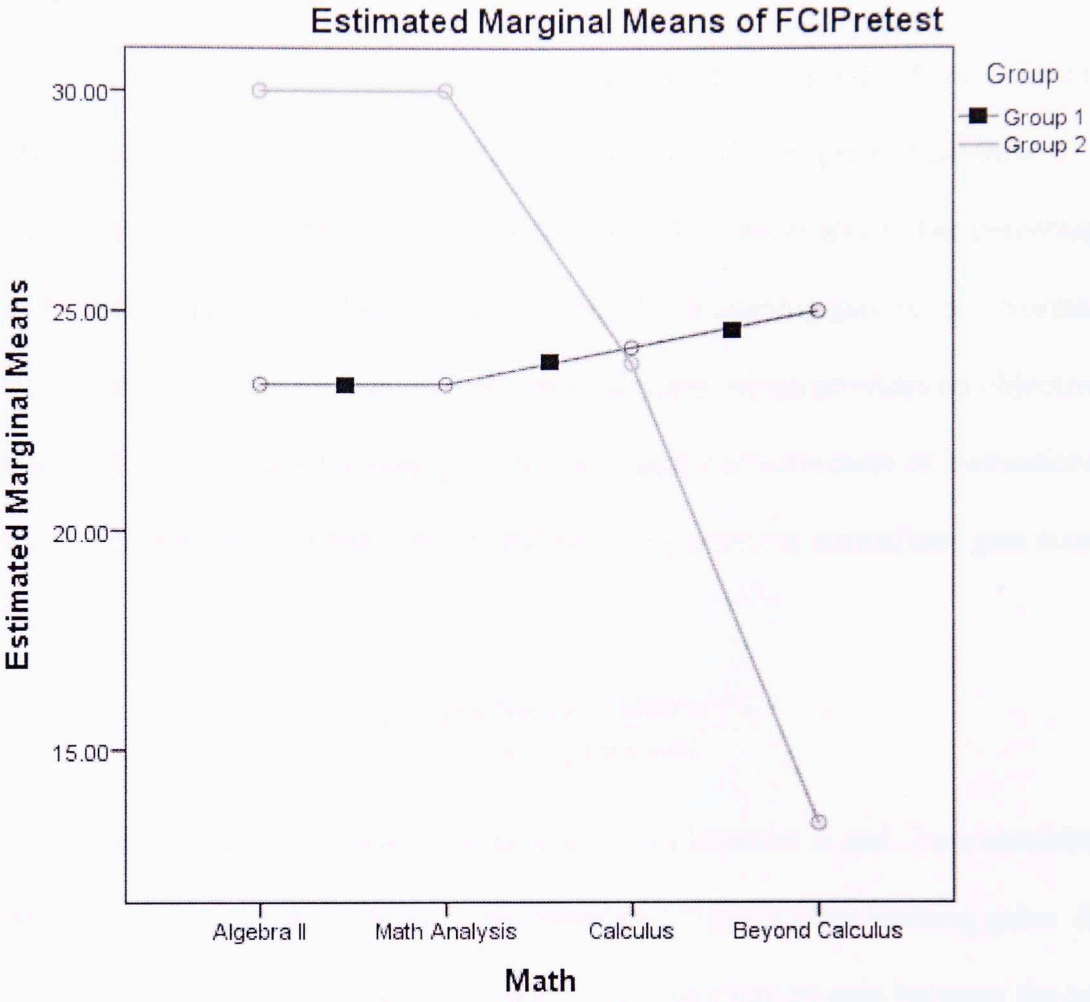


Figure 4. Plot of FCI pretest percentage scores vs. math class for Group 1 and Group 2.

Normalized Gain Scores

To answer the third research question, the FCI scores, pre-test scores from the sets of concept questions, and the post-test scores from the six parts of the three benchmark tests were turned into percentage scores for data analysis. The percentage scores from the tests were then turned into normalized learning gain scores. Normalized gain scores remove the correlation to the pre-test score, which provides an objective measure of average class learning gains to compare the effectiveness of instructional strategies (Hake, 1998). Hake used the following equation for normalized gain scores (G):

$$G = \frac{postscore\% - prescore\%}{100 - prescore\%}$$

G scores below .3 are considered low; G scores at or between .3 and .7 are considered medium; and G scores at or above .7 are considered high for class learning gains. All of the assessment pairs in the study showed a positive normalized gain between the pretest and posttest assessment, and the range of improvement was consistent for Group 1 and Group 2 between all of the Concept Questions and Benchmarks as shown in Table 10. Both groups had medium normalized learning gains for the first three pairs of assessments. Both groups had high normalized learning gains for the fourth pair of assessments, and both groups had low normalized learning gains for the fifth and sixth pair of assessments. The FCI pretest and posttest was the only pair of assessments that was in a different range for Group 1 and Group 2. Group 1 had a low normalized learning gain with a score of .282, and Group 2 had a medium normalized learning gain with a score of .345.

Table 14

Range of Normalized Gain Scores for Assessments

Group	Assessment	Average Normalized Gain	Range
1	CQ1-B1A	.314	Medium
	CQ2-B1B	.560	Medium
	CQ3-B2A	.539	Medium
	CQ4-B2B	.775	High
	CQ5-B3A	.273	Low
	CQ6-B3B	.107	Low
	FCI_{post}-FCI_{pre}	.282	Low
2	CQ1-B1A	.457	Medium
	CQ2-B1B	.538	Medium
	CQ3-B2A	.684	Medium
	CQ4-B2B	.780	High
	CQ5-B3A	.161	Low
	CQ6-B3B	.245	Low
	FCI_{post}-FCI_{pre}	.345	Medium

Note. CQ = Concept Questions; B = Benchmark; FCI = Force Concept Inventory; Range refers to Hake's (1998) categorization of the size of normalized gain scores.

Relationship of Gender, GPA and Math Level to FCI Normalized Gain Scores

A one-way between-subjects ANOVA was conducted to compare the effect of gender on FCI normalized gain scores. The within-subjects effect referred to the change in FCI normalized gain scores within each separate gender, and the between - subjects effect referred to the difference between the average male and average female FCI normalized gain score. Gender had a significant effect on FCI normalized gain scores for Group 1 as measured between males ($M = .329$, $SD = .157$) and females ($M = .216$, $SD = .212$); $F(1, 41) = 4.10$, $p = .049$. Gender had no significant effect on FCI normalized gain scores for Group 2 as measured between males ($M = .340$, $SD =.194$) and females ($M = .350$, $SD = .183$); $F(1, 26) = .019$, $p = .891$. Males scored significantly higher on the FCI normalized gain scores than females in Group 1. The descriptive statistics of gender on FCI normalized gain scores for Group 1and Group 2 are in Table 15, and the ANOVA results for the effects of gender on FCI normalized gain scores for Group 1and Group 2 are in Table 16.

Table 15

Descriptive Statistics of Gender on FCI Normalized Gain Scores for Group 1and Group 2

Group	Gender	*n*	*M*	*SD*
1	Male	25	.329	.156
	Female	18	.216	.212
	Total	43	.281	.189
2	Male	16	.340	.194
	Female	12	.350	.183
	Total	28	.344	.186

Table 16

One-Way ANOVA of Gender Effect on FCI Normalized Gain Scores for Group 1 and Group 2

Group	Effects	*SS*	df	*MS*	*F*	*p*
1	Between Group	.135	1	.135	4.10	.049
	Within Group	1.35	41	.033		
	Total	1.49	42			
2	Between Group	.001	1	.001	.019	.891
	Within Group	.934	26	.036		
	Total	.935	27			

The one-way between-subjects effect ANOVA test shows that there is no significant effect of gender on the FCI normalized gain for Group 2. There is a significant effect of gender on the FCI normalized gain for Group 1, and the males in Group 2 score significantly higher than the females in Group 1.

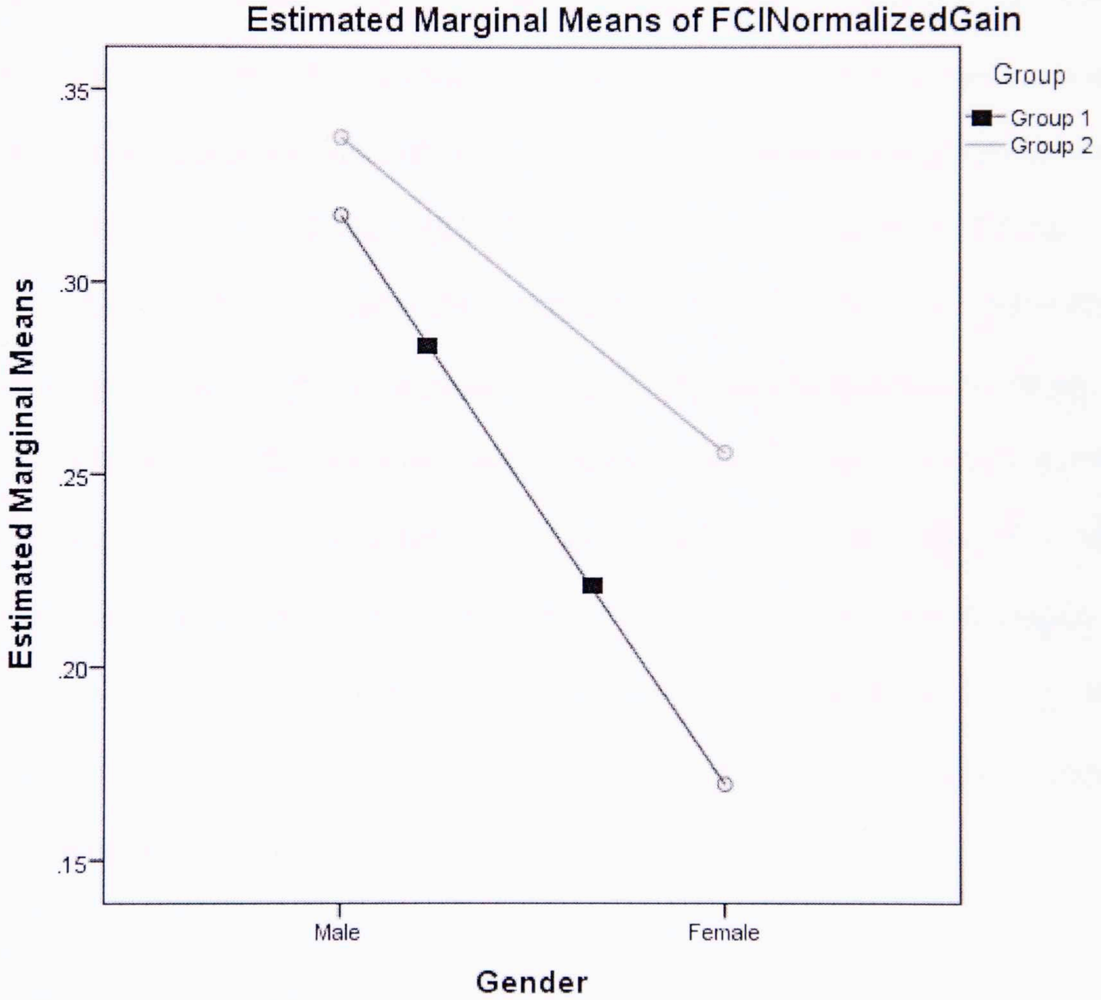


Figure 5. Plot of FCI normalized gain scores vs. gender for Group 1 and Group 2.

A one-way between-subjects ANOVA was conducted to compare the effect of GPA on FCI normalized gain scores. GPA was classified into five categories. Category 1 GPA ranged from 4.0-3.75, Category 2 ranged from 3.74-3.5, Category 3 ranged from 3.49-3.25, Category 4 ranged from 3.24-3.0, and Category 5 GPA was less than 3.0. The within-subjects effect referred to the change in FCI normalized gain scores within each separate GPA, and the between subjects effect referred to the differences between the categories of GPA on the FCI normalized gain score. GPA had no significant effect on FCI normalized gain scores for Group 1 as measured between categories, 4.0-3.75 (*M* = .298, *SD* = .154), 3.74-3.5 (*M* = .330, *SD* = .257), 3.49-3.25 (*M* = .102, *SD* = .161), 3.24-3.0 (*M* = .190, *SD* = .231), and less than 3.0 (*M* = .453, *SD* = .273); *F*(4, 38) = 2.32, *p* = .074. GPA had no significant effect on FCI normalized gain scores for Group 2 as measured between categories, 4.0-3.75 (*M* = .371, *SD* = .171), 3.74-3.5 (*M* = .403, *SD* = .025), 3.49-3.25 (*M* = .265, *SD* = .314), 3.24-3.0 (*M* = .306, *SD* = .163), and less than 3.0 (*M* = .324, *SD* = .211); *F*(4, 23) = .336, *p* = .851. The descriptive statistics of GPA on FCI normalized gain scores for Group 1and Group 2 are in Table 17, and the ANOVA results for the effects of GPA on FCI normalized gain scores for Group 1and Group 2 are in Table 18.

Table 17

Descriptive Statistics of GPA on FCI Normalized Gain Scores for Group 1and Group 2

Group	GPA	*n*	*M*	*SD*
1	4.0-3.75	28	.298	.154
	3.74-3.5	4	.330	.257
	3.49-3.25	5	.102	.161
	3.24-3.0	3	.190	.231
	< 3.0	3	.453	.273
	Total	43	.282	.188
2	4.0-3.75	13	.371	.171
	3.74-3.5	3	.403	.025
	3.49-3.25	4	.265	.314
	3.24-3.0	3	.306	.163
	< 3.0	5	.324	.211
	Total	28	.344	.186

Table 18

One-Way ANOVA of GPA Effect on FCI Normalized Gain Scores for Group 1and Group 2

Group	Effects	*SS*	df	*MS*	*F*	*p*
1	Between Group	.292	4	.073	2.32	.074
	Within Group	1.19	38	.031		
	Total	1.48	42			
2	Between Group	.052	4	.013	.336	.851
	Within Group	.883	23	.038		
	Total	.935	27			

The one-way between-subjects effect ANOVA test shows that there is no significant effect of GPA on the FCI normalized gain for Group 1 or Group 2.

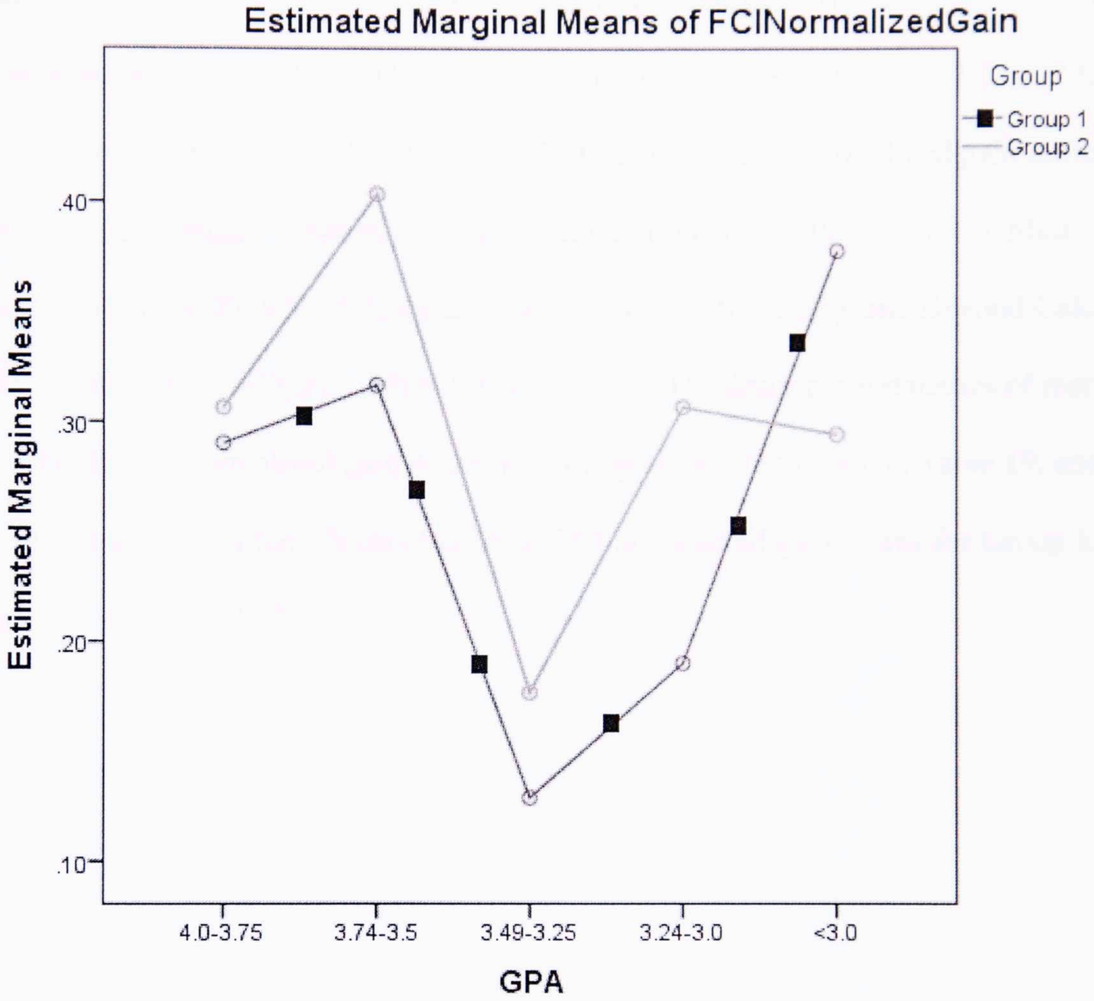


Figure 6. Plot of FCI normalized gain scores vs. GPA for Group 1 and Group 2.

A one-way between-subjects ANOVA was conducted to compare the effect of math level on FCI normalized scores. A student's math level was classified into four categories based on their current math class. Students placed in Category 1 were currently in Algebra II, Category 2 students were currently in Math Analysis, Category 3 students were currently in Calculus, and Category 4 students had already completed Calculus. The within-subjects effect referred to the change in FCI pretest scores within each separate math level, and the between-subjects effect referred to the differences between the categories of math level on the FCI normalized gain score. Math level had no significant effect on FCI normalized gain scores for Group 1 as measured between categories, Algebra II (*M* = 7.00, *SD* = 4.24), Math Analysis (*M* = 6.94, *SD* = 2.43), Calculus (*M* = 7.11, *SD* = 3.02), and Beyond Calculus (*M* = 7.50, *SD* = 2.17); *F*(3, 39) = .065, *p* = .980. Math level had no significant effect on FCI normalized gain scores for Group 2 as measured between categories, Algebra II (*M* = 9.00, *SD* = 4.42), Math Analysis (*M* = 9.00, *SD* = 4.12), Calculus (*M* = 7.15, *SD* = 2.91), and Beyond Calculus (*M* = 4.00, *SD* = 1.41); *F*(3, 24) = 1.454, *p* = .252. The descriptive statistics of math level with FCI normalized gain scores for Group 1and Group 2 are in Table 19, and the ANOVA results for the effects of GPA on FCI normalized gain scores for Group 1and Group 2 are in Table 20.

Table 19

Descriptive Statistics of Math Level with FCI Normalized Gain Scores for Group 1and Group 2

Group	Math Level	*n*	*M*	*SD*
1	Algebra II	2	.210	.085
	Math Analysis	16	.328	.213
	Calculus	19	.235	.173
	Beyond Calculus	6	.332	.177
	Total	43	.282	.188
2	Algebra II	2	.145	.205
	Math Analysis	11	.316	.217
	Calculus	13	.381	.156
	Beyond Calculus	2	.460	.028
	Total	28	.344	.186

Table 20

One-Way ANOVA of Math Level Effect on FCI Normalized Gain Scores for Group 1and Group 2

Group	Effects	*SS*	df	*MS*	*F*	*p*
1	Between Group	.101	3	.034	.944	.429
	Within Group	1.39	39	.036		
	Total	1.49	42			
2	Between Group	.132	3	.044	1.32	.292
	Within Group	.803	24	.033		
	Total	.935	27			

The one-way between-subjects effect ANOVA test shows that there is no significant effect of math level on the FCI normalized gain for Group 1 or Group 2.

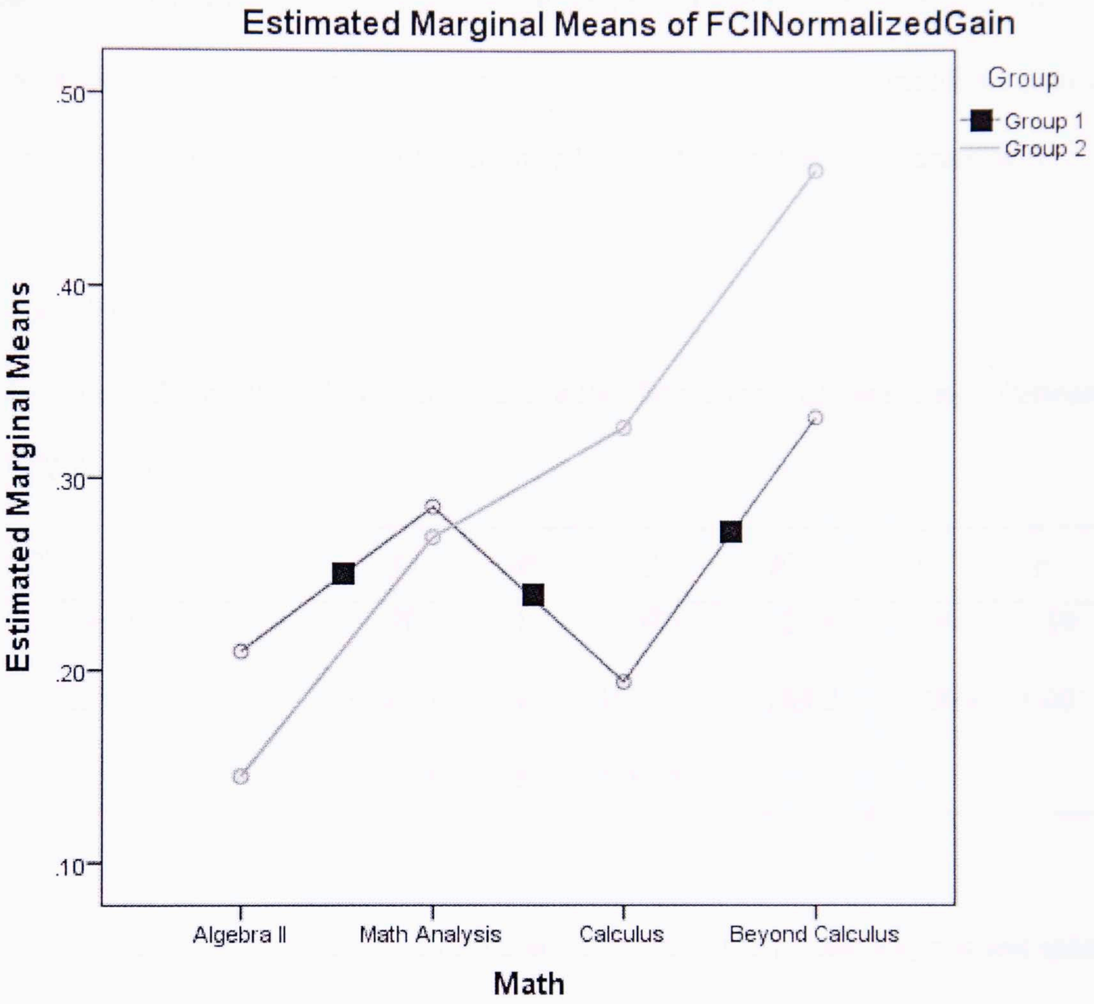


Figure 7. Plot of FCI normalized gain scores vs. math class for Group 1 and Group 2.

Repeated Measures

A repeated measures GLM test was used to determine if there was any significant difference between Group 1 and Group 2 normalized gain score for the FCI (see Table 21). The mean difference for the within-subject effects represents the change between the pretest and posttest score, and the mean difference for the between-subject effect represents the change between Group 1 and Group 2. In this case the mean difference score within-subject effect equaled 23.576, and it was significant; $F(1) = 189.2$, $p = .000$. The mean difference score for between-subject effects equaled 4.567, and it was not significant in this case; $F(1) = 189.2$, $p = .094$. The results indicate that there was a significant difference between the pretest and posttest scores, but that there was no significant difference between Group 1 and Group 2 FCI normalized gain even though Table 10 ranked Group 1 as a low gain and Group 2 as a moderate gain.

Table 21

Repeated Measures GLM Test to Measure the Significance of the Change Between FCI Scores

Effects	*SS*	*df*	*MS*	*F*	*p*	P	*eta*
Between Group	707.3	1	707.3	2.88	.094	.39	.036
Within Group	18850.5	1	1885.5	189.2	.000	1.00	
Total	19557.8	2	195578				

Figure 8 shows the significant change between the pretest and posttest scores for each group and the non-significant change between Group 1 and Group 2 according to the repeated measures test.

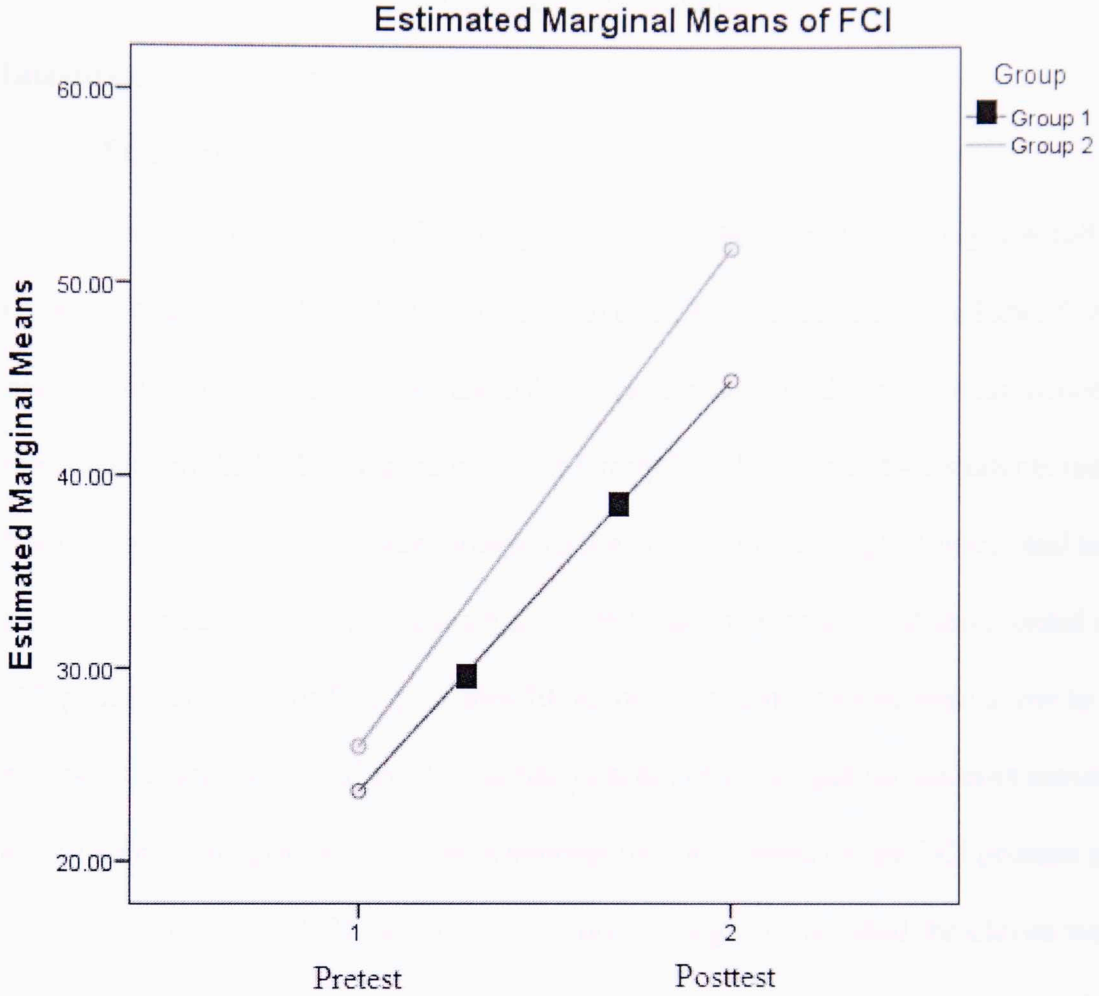


Figure 8. Plot of FCI percentage scores on the pretest and posttest for Group 1 and Group 2.

Chapter V-Discussion

Interpretation of Data

FCI scores.

Overall, the scores for the FCI pretest and posttest were relatively low following Hestenes and Halloun's (1995) interpretation of FCI scores as shown in Table 5. All students in Group 1 and Group 2 had a *low* understanding of the Newtonian concept of force based on the FCI pretest scores. For Group 1, only 10 out of 43 students moved from a *low* to a *moderate* understanding of the Newtonian concept of force, and no students moved to a *high* understanding of the Newtonian concept of force based on the FCI posttest scores. For Group 2, only 10 out of 28 students moved from a *low* to a *moderate* understanding of the Newtonian concept of force, and no students moved to a *high* understanding of the Newtonian concept of force based on the FCI posttest scores. According to Hake (1998), this means that no one began or finished the eleven week study with a completely consistent view of the Newtonian concept of force based on the FCI pretest and posttest scores of the participants.

There are other factors besides FCI percentage scores to look at when evaluating students' progress towards a complete conceptual understanding of the Newtonian concept of force. Hestenes and Halloun (1995) emphasized that the FCI should not be given as a test based on percentages because the entire test was designed to confront students with common misconceptions to show the subtle differences of how to interpret and apply Newton's laws correctly in different situations. The FCI is designed to test multiple concepts about force, which means that the learning process can be seen as a process over time rather than a onetime event where students either have no knowledge or a complete understanding of the concept. The FCI should be used to gauge the understanding of a student about force and to gauge if a student is progressing in their understanding of force over time.

The goal of the high school physics class used in this study was not to replace a university level physics class. The goal was to prepare students to succeed in a college physics class. Lasry, Mazur, and Watkins (2008) found that students with low background information indicated by low FCI pretest scores had statistically significant lower normalized learning gains on the FCI than students with more background knowledge indicated by high FCI pretest scores. Taking physics in high school gives students additional time to develop a framework of how the major physics concepts work together to explain how the world works. When they see the information again in college, all of the new information can be weighed against this past understanding. Their old system is probably not completely consistent with the Newtonian concept of force, but it gives them a good starting place to reevaluate how the world works.

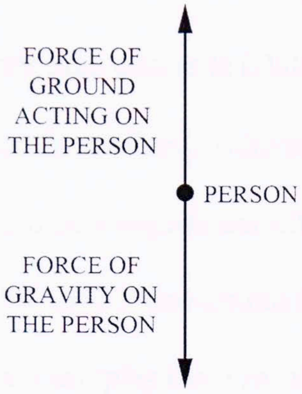
The compressed timeframe of the study also required that the FCI posttest be given before the class had discussed friction, inclines, pulleys, or centripetal force. It is likely that the FCI scores would have been higher if students would have been given the opportunity to apply Newton’s laws in these new situations. A benefit of administering the FCI while students were still experiencing the unit was that the FCI was used after the study as a learning tool to help students confront their misconceptions. According to Marek (2009), part of educating students involves presenting the students with situations or thought experiments that will cause them to disequilibrate to create the opportunity for the students to create a new mental structure to accommodate the new information.

The abstract nature of physics concepts and the difficulty of synthesizing the concepts in a real situation could be contributing factors to the *low* scores on the FCI. Many of the students are transitioning from being concrete thinkers to being formal thinkers. To this end, most all of the students are capable of reciting Newton's laws, but few students may have developed the level of critical thought that is necessary to develop a consistent understanding of how Newton's laws are applied. Since all of the Newtonian concepts are interconnected, students must synthesize these concepts to understand fully what is happening in a given situation. Figure 9 is an example of the different ideas that students are applying when they look at one example of a concept question on forces.

Question

Are the forces acting on a person standing on the ground Newton's 3rd law forces?

Diagrams

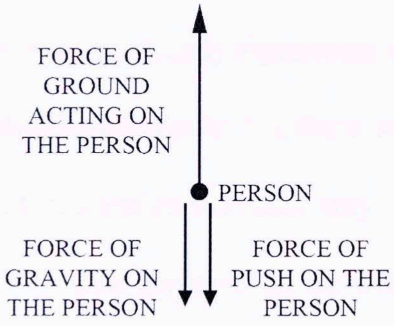


Statement 1:

The forces on the person are equal and opposite because there is no acceleration.

Statement 2:

This is a consequence of Newton's 2nd law that states that the acceleration of an object is the result of the sum of all of the forces acting on the object.

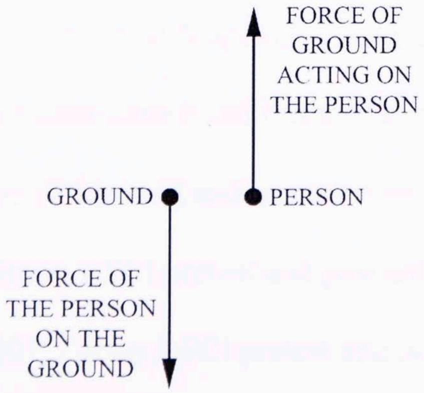


Statement 3:

The forces could be unequal, if a push was applied down on the person.

Statement 4:

Newton's 3rd law forces are always equal and opposite by definition.



Statement 5:

An example of a 3rd law force in the situation is the force of the person acting on the ground (action), and the force of the ground acting on the person (reaction).

Conclusion:

The forces acting on the person are not 3rd law forces.

Figure 9. Example of concepts involved with correctly applying Newton's laws.

Paired samples t-test.

The first two hypotheses were that both peer-led and teacher-led discussions would cause statistically significant learning gains because both treatments were tested against physics students' prior knowledge, which has been shown to contain non-Newtonian concepts (Halloun & Hestenes, 1985). In this study, both the first and second hypothesis were supported for the overall classes since there was a positive, statistically significant difference between the pretest scores and the posttest scores for both Group 1 and Group 2 for all of the assessments based on the significance of the paired samples t-test results in Table 6.

The correlation of the paired samples t-test results in Table 7 showed that there was no statistically significant correlation between the pretest and posttest scores for individual students. So, there was improvement, but all students were not affected by the treatment in the same way.

One explanation for why there was no correlation between the pretest and posttest scores could be that there were not enough questions to clearly gauge the individual student's conceptual understanding. The FCI pre- and posttest, which contains over 30 questions, were the only pairs of statistically significant correlations between pretest and posttest scores for both Group 1 and Group 2. The correlation for the FCI pretest and posttest was also the highest out of all of the assessment pairs. Group 1 FCI pretest and posttest had a *moderate* correlation of .491 and a *p*-value of .001. Group 2 FCI pretest and posttest had a moderate correlation of .444 and a *p*-value of .018. Refer to Table 7 for all correlation values. The concept questions and benchmarks had fewer questions to gauge students understanding than the FCI, and so the inclusion of more questions in the concept questions and benchmarks could increase the probability of finding statistically significant correlations between the pretest and posttest scores for the concept questions and benchmarks.

Another explanation of the lack of a statistically significant correlation between the pretest and posttest scores was the inclusion of other activities and time between administering the pretests and posttests. The data suggest that there was an increase in scores between pretest and posttest scores, but this increase could have been due to many other factors other than the treatment. Only 10%-20% of the class time was dedicated to the conceptual tests. The classes also participated in labs, discussions, problems, and homework that could have caused the learning gain for an individual student. Minimizing these other factors would increase the correlation between the pretest and posttest scores.

Relationship of gender, GPA, and math level to FCI pretest.

Gender was the only significant independent variable to affect the FCI pretest scores for Group 1, and none of the variables significantly affected the FCI pretest scores for Group 2. GPA and math level were not significant predictors of FCI pretest scores. As previously discussed, the FCI is a challenging test that confronts common assumptions, which could be why students with high GPAs and more math classes did not produce higher FCI pretest scores. Gender does have a significant effect on FCI pretest scores for Group 1. Lorenzo, Crouch, and Mazur (2006) confirmed that males typically score higher on the FCI, but there is a debate over whether this difference is from gender bias in the FCI or if it is showing that males typically understand the concept of force better than females. As a classroom teacher, the principal investigator has not observed that males possess a greater understanding of the concept of force than females.

Normalized gain scores.

Overall, the normalized gain scores were *low* to *medium* for students in both groups. However, there was a positive learning gain between the pretest and posttest scores for all of the assessments, which shows that students were in the process of gaining a better understanding of the Newtonian concept of force. The objective is to have high normalized learning gains, but the same arguments discussed earlier in chapter 5 under the FCI scores section about the difficulty of the questions, the process of learning, and the goals of the high school physics class could also be applied to the normalized gain scores.

All of the normalized gain scores are listed in Table 14. Both groups had *medium* normalized gain scores on the first, second, and third sets of concept questions. Both groups had *high* normalized gain scores on the fifth set of concept questions. Both groups had *low* normalized gains on the fifth and sixth sets of concept questions. The consistent range of normalized gain scores between the concept questions and benchmarks indicated that both treatments had similar effects. However, this finding is limited by the lack of correlation between the pretest and posttest scores. The study cannot support that the treatment is what caused the change, and so it cannot be said that the two, distinct discussion formats had the same effect on conceptual understanding.

There was a different range for the FCI normalized gain score for Group 1 and Group 2. The FCI normalized gain score for Group 1 was .2818, which is a *low* gain score, and the FCI normalized gain score for Group 2 was .3452, which is a *medium* gain score. The repeated measures test will be discussed later to show if there was a significant difference between the FCI normalized gain scores for Group 1 and Group 2.

Relationship of gender, GPA and math level to FCI normalized gain score.

Gender was the only significant independent variable to affect the FCI pretest scores for Group 2, and none of the variables significantly affected the FCI pretest scores for Group 1. GPA and math level were again not significant predictors of FCI normalized gain scores. As discussed above, the FCI is a challenging test that confronts common assumptions, which could be why students with high GPAs and more math classes did not produce higher FCI pretest scores. Gender does have a significant effect on FCI normalized gain scores for Group 2; males scored higher than females. These findings agree with Lorenzo, Crouch, and Mazur (2006) in that males typically scored higher than females on the FCI. Treatment 1, using peer-led discussions, did have a more positive effect on females than males so that the gender gap was no longer statistically significant on the normalized learning gain. Treatment 2, using teacher-led discussions, did have a more positive effect on males than females so that the gender gap became significant on the normalized learning gain. Most females may have more significant learning gains than males from using peer discussions, while most males may have more significant learning gains than females from individually answering concept questions.

Repeated measures.

The repeated measures test showed that there was a significant change within-groups between the FCI pretest and posttest scores for Group 1 and Group 2; $F(1) =$

189.2, $p = .000$. The repeated measures test also showed that there was no significant change between Group 1 and Group 2 FCI pretest and posttest scores; $F(1) = 2.880$, p .094. This meant that the final hypothesis concerning the comparison between the peer-led and teacher-led discussions was not supported by the data. Students who participated in the peer-led discussion approach did not show significantly greater learning gains than students in the teacher-led class discussion group as evidenced by the lack of significance in the between measures test for the FCI pretest and posttest. Even if there had been significant change between the groups, the result would still have been limited by the lack of correlation between the assessment pairs, which would have limited the ability to conclude that any change would have been caused by the treatment method.

Observations from the Principal Investigator as the Classroom Teacher

The third hypothesis that there will be a significant difference between the two treatment methods was not supported by the data. But, the lack of correlation between the FCI pretest and posttest signifies that the scores from the assessments may not accurately measure the effect of the treatment. From observations in both treatment classes, the principal investigator would argue that Group 1 treatment was more effective than Group 2 treatment. It appeared that everyone participated in the Group 1 discussions after answering the questions individually, and students seemed engaged in defending their point of view or asking for clarification. However, Group 2 treatment did not spark the same participation as Group 1 treatment. Everyone answered the first round of clicker questions individually in Group 2, but there was more negative feedback in the justifications.

One problem with Group 2 treatment was that the questions were specifically chosen to challenge the students and spark discussions. But, students in Group 2 were asked to answer and justify the question individually. Rather than sparking discussions, the questions caused some of the students to become frustrated because they felt that they could not conceptualize the correct answer. Student frustration led some students to answer that they did not know, that they were stupid, or to put a one word answer with no explanation rather than writing out a justification for their answer. The principal investigator attempted to model writing out true statements about the problem to show how to analyze the problem and justify an answer, but this did not seem to help students reason through the new concepts.

Another limitation of Group 2 treatment was that it took more time to administer each question. Group 1 peer-led discussion elicited students' ideas. Therefore, only a quick summary was needed to complete the question-centered discussion. Even if the majority did not have the Newtonian answer, the teacher quickly addressed problems with the majority's choice, and other students in the room justified why their answer was consistent with the Newtonian concept. With Group 2 treatment, the instructor collected and summarized all of the responses before again asking the students to choose a response and discuss the choices with the whole class. The instructor tried writing selected responses on the whiteboard, reading selected responses, and putting responses under the document camera. However, this method was always more hurried for time than Group 1.

Limitations to Study

The compressed timeframe for the study during a twelve-week period was a limitation, and it may not have been adequate time to develop a statistically significant difference between Group 1 and Group 2. During the administration of the six sets of concept questions was the only time when the classes experienced the specific treatment method. The requirements of finishing the thesis also mandated that the FCI posttest was given before the classes had covered friction, incline planes, pulleys, and centripetal force. So, the students had covered the basics of Newton's three laws, but did not have time to expand the laws into other applications to help develop a broader understanding of the concepts that also could have helped the students perform better on the FCI.

Another limitation to the study was that the principal investigator was also the classroom teacher. The assent and consent forms were collected by a third party so that the principal investigator was not aware of who was participating in the study until after all of the data were collected. This was to minimize any potential coercion to students, but the principal investigator still taught the students, which meant that there was the potential to influence the research. The investigator believed that he tried to error on the side of being the classroom teacher rather than a researcher. For example, the whole research schedule was arranged around the different activities of the class, and time was taken away from the concept questions if it appeared that doing a different activity would be more productive at that time. In addition, restricting students' access to the concept questions after they had answered them in class could have caused the benchmarks to more closely relate to what the students learned from the discussions.

However, the principal investigator was more concerned with the students learning the material, which meant that he encouraged the students to review the concept questions online before completing the benchmarks.

The study itself was also limited by the fact that it was only conducted by one teacher at one high school. This limits the ability of the study to be generalized to other teachers at other schools, but the results can be used to help structure a larger scale study involving multiple schools and multiple teachers based on what was learned from conducting the study.

Recommendations for Future Research

The study was primarily a learning experience in which the principal investigator learned about the factors involved in collecting data within the classroom. There was a significant increase between the pretest and posttest scores for each pair of assessments, which is worth studying in the future. The Group 1 peer-led discussions seemed to be very effective to the principal investigator, but changes would need to be made to Group 2. One modification to Group 2 would be to ask more direct questions for the pretest so that students would be able to practice applying the concepts in their justifications without the same level of frustration. Another modification that would need to be made to Group 2 would be to find a more efficient way of disseminating the student responses to the class to more efficiently use the time in the teacher-led discussion.

The most significant design flaw in the study was the amount of time between administering the pretest and posttest assessments. The data suggests that there was an increase in scores between pretest and posttest scores, but there was not a large correlation between the pretest and posttest scores. To address this problem, concept questions could be given at the end of the unit as a review, and the benchmark would be given the very next day to increase the correlation by removing the other factors between testing. Another possibility would be to give the concept questions within the unit but to pair each individual question with a benchmark question. The students would go through the treatment with one concept question, and then immediately answer the paired benchmark question to gauge the effect of the treatment. In addition to removing the other factors influencing the scores, this would also increase the correlation because each pair of questions would focus on an individual concept rather than being a combination of questions over multiple concepts from an entire section. Students would also be motivated to discuss the question if the initial question was graded on participation while the follow-up question over the same concept would be graded on correctness.

It is recommended that the FCI still be given as a pretest and posttest for any future research so that the research can be compared to other studies. The FCI also serves as a good cumulative evaluation of the Newtonian concept of force. The time between the pretest and posttest is not as much of a concern as it is with the smaller pairs of assessments because it is intended to be a cumulative assessment of how an individual's systematic understanding of force has changed over the length of the study.

Conclusion

The study showed a significant difference between the pretest and posttest scores for all of the assessments, but the lack of a strong correlation between assessments revealed problems with the design of the study that prevented an accurate gauge of the treatment effect on each group. These problems included the time between the pretests and posttests for both groups and the lack of a time efficient method of representing the student responses for discussions for Group 2. The data did not support the hypothesis that there was a statistically significant difference between the FCI pretest and posttest scores for Group 1 and Group 2. The principal investigator believes that the peer-led discussions in Group 1 were more effective than the teacher-led discussions based on the productive discussions observed in Group 1 and the frustration observed in Group 2. The peer-led discussions within the Peer Instruction method promoted more participation by providing the students with the opportunity to discuss the possibilities in small groups. Peer-led discussions exposed the students to many more viewpoints than the teacher-led discussions where the students had to answer the questions individually. Future research could help clarify what factors affect student learning gains, but changes would need to be made to the research design to strengthen the correlation between the pretest and posttest scores. The principal investigator believes that the implementation of programs, like Peer Instruction, in high school physics classes is critical to increasing the conceptual understanding of students. Effective classroom research will continue to improve the methods of teaching and strengthen the case for the inclusion of student centered learning within the classroom.

References

Adams, H., & Howard, L. (2009). Clever clickers: Using audience response systems in the classroom. *Library Media Connection*, *28*(2), 54-56.

Aryal, B., & Zollman, D. A. (2007). Investigating Peer Scaffolding in Learning and Transfer of Learning Using Teaching Interviews: *2007 PERC Proceedings*, AIP Conference Preceding Number 951. Melville, NY.

Athee, M. (1993). The proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics: *A survey of Finnish pupils about thermal phenomena*. Ithaca, NY: Misconceptions Trust.

Auras, R., & Bix, L. (2007). Wake up! The effectiveness of a student response system in large packaging classes. *Packaging Technology and Science*, *20*, 183-195.

Bar, V., & Travis, A. S. (1991). Children's viewpoints concerning phase changes. *Journal of Research in Science Teaching*, *28*, 363-382.

Baser, M., & Geban, O. (2008). Effect of instruction based on conceptual change activities on students' understanding of static electricity concepts. *Research in Science & Technological Education*, *25*(2), 243-267.

Beatty, I. (2004). Transforming student learning with classroom communication systems. *Educause*, *3*, 1-13.

Bligh, D. (2000). *What's the Use of Lectures?* San Francisco, Ca: Jossey-Bass.

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students, *Science Education*, *68*, 493-503.

Bybee, R. W., & Sund, R. B. (1990). *Piaget for Educator* (2nd ed.). Prospect Heights, IL: Waveland Press.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, *50*, 66–71.

Cohen, E. R., Eylon, B., & Ganiel, U. (1983). Potential difference and current in simple electric circuits: A study of students' concepts. *American Journal of Physics*, *51*, 407-412.

Crouch, C. H., & Mazur, E. (2001). PI: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970-977.

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer Instruction: Engaging students one-on-one, all at once. *Research Based Reform of University Physics*, *1*, 1-50.

Cummings, K., & Roberts, S. G. (2008). Proceedings from PER Conference Series '08: A study of Peer Instruction methods with high school physics students. *Physics Education Research Conference*, Edmonton, Canada.

DeHaan, R. L. (2005). The impending revolution in undergraduate science education. *Journal of Science Education and Technology*, *14*(2), 253-269.

Dilber, R., Karaman, I., & Duzgun, B. (2009). High school students' understanding of projectile motion concepts. *Education Research and Evaluation*, *15*, 203-222.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, *20*, 81-94.

Duit, R. (2002). *Bibliography STCSE: Students' and teacher' conceptions and science education.* Kiel, Germany: IPN – Leibniz Institute for Science Education. (http://www.ipn.uni-kiel.de/aktuell/stcse/bibint.html).

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*, 671-688.

Duncan, D. (2005). *Clickers in the classroom: How to enhance science teaching using classroom response systems*. San Francisco: Pearson/Addison Wesley.

Dupin, J., & Johsua, S. (1987). Conceptions of French pupils concerning electric circuits: Structure and evolution. *Journal of Research in Science Teaching*, *42*, 791-806.

Educational Policies Commission (1961). *The central purpose* of *American education.* Washington, D.C.: National Education Association.

Edwards, H., Smith, B. A., & Webb, G. (2001). *Lecturing: Case studies, experience and practice*. London, England: Kogan.

Ericson, G. L. (1979). Children's conceptions of heat and temperature. *Science Education*, *63*, 221-230.

Eryilmaz, A. (2002). Effects of conceptual assignments and conceptual change discussion on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching*, *39*(10), 1001–1015.

Fagen, A. P., Crouch, C. H., & Mazur, E. (2002). Peer Instruction: Results from a range of classrooms. *The Physics Teacher*, *40*, 206-209.

Feher, E., & Rice, K. (1992). Shadows and anti-images: Children's' conceptions of color. *Journal of Research in Science Teaching*, *29*, 505-520.

Fredette, N., & Lochead, J. (1980). Student conceptions of simple circuits. *The Physics Teacher*, *18*, 194-198.

Gilbert, J. K., Osborne, R. J., & Fensham, P. (1982). Children's science and its implications for teaching. *Science Education*, *66*, 625-633.

Goldberg, F. M., & McDermott, L. C. (1986). Students' difficulties in understanding image formation by a plane mirror. *The Physics Teacher*, *24*, 472-480.

Goldberg, F. M., & McDermott, L. C. (1987). An investigation of students' understanding of real image formed by a converging lens or concave mirror. *American Journal of Physics*, *55*, 108-119.

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

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

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

Hatch, J., Jensen, M., & Moore, R. (2005). Manna from heaven or clickers from hell: Experience with an electronic response system. *Journal of College Science Teaching*, *34*, 36-39.

Heller, P., & Finley, F. N. (1992). Variable uses of alternative conceptions: A case study in current electricity. *Journal of Research in Science Teaching*, *29*, 259-275.

Hestenes, D., & Halloun, I. (1995). Interpreting the Force Concept Inventory. *Physics Teacher*, *33*, 502-506.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *Physics Teacher*, *30*, 141-158.

Huffman, D., & Heller, P. (1995). What does the Force Concept Inventory actually measure? *Physics Teacher*, *33*, 138-143.

Idar, J., & Ganiel, U. (1985). Learning difficulties in high school physics: Development of a remedial teaching method and assessment of its impact on achievement. *Journal of Research in Science Teaching*, *22*, 127-140.

Iverson, H. L., Briggs, D. C., Ruiz-Primo, M. A., Talbot III, R. M., & Shepard, L. A. (2009). Proceedings from Physics Education Research Conference '09. *Undergraduate physics course innovations and their impact on student learning*. Ann Arbor, Michigan: American Institute of Physics.

James, M. C. (2006). The effect of grading incentive on student discourse in Peer Instruction. *American Journal of Physics*, *74*(8), 689-691.

Johnson, D. W., & Johnson, R. T. (1991). *Active learning: Cooperation in the College Classroom*. Edina, MN: Interaction Book Company.

Kalman, C. S., Miner-Bolotin, M., & Antimirova, T. (2010). Comparison of the effectiveness of collaborative groups and peer instruction in a large introductory physics course for science majors. *Canadian Journal of Physics*, *88*, 325-332.

Keller, C., Finkelstein, N., Perkins, K., Pollock, S., Turpen, C., & Dubson, M. (2007). Physics Education Research Conference of '07: Research-based practices for effective clicker use. Greensboro, NC: American Institute of Physics.

King, A. (1992). Facilitating elaborative learning through guided student generated questioning. *Educational Psychologist, 27,* 111-126.

Koch, D. (2009). Evaluation Instruments. Retrieved from http://modeling.asu.edu/R&E/Research.html

Kuhn, T. S. (1970). *Structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Lasry, N. (2008). Clickers or flashcards: Is there really a difference? *The Physics Teacher*, *46*(4), 242-246.

Lasry, N., Mazur, E., & Watkins, J. (2008). Peer Instruction: From Harvard to the two-year college. *American Association of Physics Teachers*, *76*(11), 1066-1069.

Latessa, R., & Mouw, D. (2005). Use of an audience response system to augment interactive learning. *Family Medicine*, *37*, 12-14.

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

Macgregor, J., Cooper, J. L., Smith, K. A. & Robinson, P. (2000). *Strategies for energizing large classes: From small groups to learning communities*. San Francisco, Ca: Jossey-Bass.

Maier, S. J. & Marek, E. A. (2005). The Learning Cycle: A Reintroduction. *The Physics Teacher, 43*, 31-35.

Maloney, D. P., O'Kuma, T. L., & Hieggelke, C. J. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics, 69,* 12-23.

Ma-Naim, C., Bar, V., & Zinn, B. (2002). Proceedings from the Third European Symposium on Conceptual Change '02: *Integrating microscopic macroscopic and energetic descriptions for a conceptual change in thermodynamics.* Turku, Finland.

Marek, E. A. (2009). Genesis and Evolution of the Learning Cycle. In W. Roth & K. Tobin (Eds.), The World of Science Education: *Handbook of Research in North America* (141-156). Sense Publishers.

Marek, E. A., & Cavallo, A. M. L. (1997). *The learning cycle: Elementary school science and beyond* (Rev. ed.). Portsmouth, NH: Heinemann.

Martyn, M. (2007). Clickers in the classroom: An active learning approach. *Educause Quarterly*, 71-74.

Marzano, R. J. (2009). Setting the record straight on "High-Yield" strategies. *Kapplan, 91*, 30-37.

Mazur, E. (1997). *Peer Instruction: A user's manual.* Upper Saddle River, NJ: Prentice Hall.

McDermott, C. (1984). Research of conceptual understanding in mechanics, *Physics Today*, *37*(7), 24-32.

McDermott, C. (1990). What we teach and what is learned: Closing the gap. *American Journal of Physics*, *59*, 301-315.

Minstrell, J. (1982). Explaining ''at rest'' condition of an object. *The Physics Teacher*, *20*, 10–14.

National Center for Education Statistics. (2011). *Program for International Student Assessement (PISA) 2011.* Retrieved from http://nces.ed.gov/surveys/pisa/.

National Center for Education Statistics. (2011). *The Nation's Report Card: Science 2009.* Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2009/2011451.asp

National Science Board. (2010). *Science and Engineering Indicators: 2010*. Retrieved from http://www.nsf.gov/statistics/seind10/c1/c1h.htm.

Nicol, D. J., & Boyle, J. T. (2003). Peer Instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom. *Studies in Higher Education*, *28*(4), 457-473.

Posner, G.J., Strike, K.A., Hewson, P.W., & Gertzog, W.A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211–227.

Powell, K. C. & Cody, K. J. (2009). Cognitive and social constructivism: Developing tools for an effective i instruction classroom. *Education, 130*, 241-250.

Rosenshine, B., Meister, C, & Chapman, S. (1996). Teaching student teachers to generate questions: A review of the intervention studies. *Review of Educational Research, 66,* 121-181.

Savinainen, A., & Scott, P. (2002). The Force Concept Inventory: A tool for monitoring student learning. *Physics Education*, *37*, 45-52.

Sencar, S., & Eryilmaz, A. (2004). Factors mediating the effect of gender on ninth-grade Turkish students' misconceptions concerning electric circuits. *Journal of Research in Science Teaching*, *41*(6), 603-616.

Saul, J. M. (1998). *Beyond problem solving: Evaluating introductory physics courses through the hidden* curriculum. Retrieved from http://www.physics.umd.edu/perg/dissertations/Saul/

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

Shayer, M., & Wyllam, H. (1981). The development of the concepts of heat and temperature in 10-13 year-olds. *Journal of Research in Science Teaching*, *18*, 419-434.

Singh, C. (2005). Impact of peer interaction on conceptual test performance. *American Journal of Physics, 73*(5), 446-451.

Smith, M., Wood, W., Krauter, K., & Knight, J. (2011). Combining peer discussion with instructor explanation Increases student learning from in-class concept questions. *Life Science Education*, *10*, 55-63.

STEM Education Coalition. (2011). retrieved from http://www.stemedcoalition.org/wp-content/uploads/2011/02/STEM-1pg.pdf

Towbridge, D.E., & McDermott, L.C. (1980). Investigation of student understanding of concept of velocity in one dimension. *American Journal of Physics*, *48*, 1020–1028.

Towbridge, D.E., & McDermott, L.C. (1981). Investigation of student understanding of concept of acceleration in one dimension. *American Journal of Physics*, *49*, 242–253.

Vionnet, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*, *1*, 205–221.
Watts, M., & de Jesus, H. P. (2005). The cause and affect of asking questions:

Reflective case studies from undergraduate sciences. *Canadian Journal of Science, Mathematics, & Technology Education*, 5(4), 437-452.

West, L. H. T., & Pines, A. L. (1985). *Cognitive structure and conceptual change*. New York, NY: Academic Press.

Willoughby, S. D. & Gustafson, E. (2009). Technology talks: Clickers and grading incentive in the large lecture hall. *American Association of Physics Teachers*, *77*(2), 180-183.

Wit, E. (2003). The use of a personal response system in statistics teaching. *MSOR Connections*, *3*, 5-11.

Wood, W. B. (2004). Clickers: A teaching gimmick that works. *Developmental Cell*, *7*, 796-798.

Yourstone, S. A., Kraye, H. S., & Albaum, G. (2008). Classroom questioning with immediate electronic response: Do clickers improve learning? *Decision Sciences Journal of Innovative Education*, *6*(1), 75-88.

Zahorik, J. A. (1996). Elementary and secondary teachers' reports of how they make learning interesting. *Elementary School Journal*, *98*, 3-13.

Appendix

Appendix A: Research Timeline

Appendix B: Consent/Assent Forms

Appendix A

Research Timeline

W	8-24	Syllabus, assent and consent forms, physics binders, and Walking and Running Lab
R	8-25	Walking and Running Lab
F	8-26	Walking and Running Lab
M	8-29	**Force Concept Inventory Test as a Pretest (Clicker Questions)**, Collect syllabus, assent and consent forms, and check for physics folders.
T	8-30	Average Velocity Over Multiple Intervals
W	8-31	Average Velocity Over Multiple Intervals
R	9-1	Converting Between Graphs
F	9-2	Graph Matching Lab, and Miniquiz over dimensional analysis
M	9-5	Holiday
T	9-6	Graph Matching Lab, front side of More Practice with Slope and Area Graphs for homework due Thursday, discuss diluting gravity with Galileo Lab.
W	9-7	Collect Data for Diluting Gravity with Galileo Lab and make graphs
R	9-8	Discuss Diluting Gravity with Galileo Lab and review for Quiz 1 over dimensional analysis, position vs. time graphs, and velocity vs. time graphs. Work back side of More Practice with Slope and Area Graphs.
F	9-9	Quiz 1
M	9-12	Finding the Relationship Between Graphs
T	9-13	**Graphing 1-D Motion (Clicker Questions)**, Finding the Relationship Between Graphs
W	9-14	Finding Relationship Between Graphs, Birth of the Kinematic (Motion) Equations
R	9-15	Review Derivations from Birth of the Kinematic Equations, One-Dimensional Motion
F	9-16	One-Dimensional Motion 2 (Finish for HW)
M	9-19	One-Dimensional Motion 3, Hand out Packet 1 cover sheet
T	9-20	One-Dimensional Motion 3, Free Fall Lab Inside
W	9-21	Free Fall Lab Outside and Discussion
R	9-22	Review for Test 1 over dimensional analysis, graphing, and one-dimensional motion equations, and how to put packet together
F	9-23	Test 1 and Packet 1 (all worksheets due)
M	9-26	**Free Fall Problems 1-D (Clicker Questions)**, Photogate demonstration for Free Fall, Objects in Free Fall
T	9-27	Objects in Free Fall
W	9-28	Graphing Free Fall Problems
R	9-29	Graphing Free Fall Problems, Reading Bodies in Motion
F	9-30	More Free Fall Problems

M	10-3	More Free Fall Problems
T	10-4	**Benchmark 1 (Clicker Questions)** and Quiz 2
W	10-5	Horizontal Projectile Motion Demonstration, Horizontal Projectile Motion
R	10-6	Fall Break
F	10-7	Fall Break
M	10-10	**Horizontal Projectile Motion (Clicker Questions)**, Horizontal Projectile Motion Lab
T	10-11	Vector Exploration
W	10-12	Vector Exploration
R	10-13	Ten Steps to a "Proper Resultant"
F	10-14	Ten Steps to a "Proper Resultant," Projectile Motion with an Angle, Demonstration, Projectile Motion with an Angle
M	10-17	**Projectile Motion with an angle (Clicker Questions),** Projectile Motion with an Angle
T	10-18	Graphing Horizontal Projectile Motion and Projectile Motion with an Angle
W	10-19	Projectile Motion with an Angle Lab
R	10-20	Projectile Motion with an Angle Lab, Projectile Motion Problems
F	10-21	Projectile Motion Problems, Projectile Motion Reading and questions
M	10-24	**Benchmark 2 (Clicker Questions)**, Inertial Balance Lab
T	10-25	Test 2
W	10-26	Force and Acceleration Lab
R	10-27	Newton's Three Laws, Third Law Computer Demonstration
F	10-28	**Newton's Laws 1-D (Clicker Questions)**, ∑F = ma finish for HW
M	10-31	Professional Day
T	11-1	∑F = ma & angles, False Statements Regarding Newton's Laws
W	11-2	∑F = ma & angles, False Statements Regarding Newton's Laws
R	11-3	**Newton's Laws 2-D (Clicker Questions),** Free Body Diagrams
F	11-4	Quiz 3
M	11-7	∑F = ma & angles 2
T	11-8	∑F = ma & angles 2
W	11-9	**Benchmark 3 (Clicker Questions),** ∑F = ma & angles 2
R	11-10	**Force Concept Inventory Post Test**
F	11-11	Parent-Teacher Conferences

Appendix B

Parent Consent and Student Assent Forms

University of Oklahoma
Institutional Review Board
Informed Consent to Participate in a Research Study

Project Title: Concept Questions with Peer-led vs. Teacher-led Discussions in a High School Physics Classroom
Principal Investigator: Kevin Warren
Department: Instructional Leadership and Academic Curriculum

Your child is being asked to volunteer for this research study. This study is being conducted at [redacted] High School. Your child was selected as a possible participant because he or she is in Mr. Warren's physics class this semester.

Please read this form and ask any questions that you may have before agreeing to allow your child to take part in this study.

Purpose of the Research Study

The purpose of this study is to compare a class with peer-led discussions versus a class with teacher-led discussions. The study will compare the learning gains for students in each group on the Force Concept Inventory and the benchmark tests. Both classes will have the same labs, worksheets, and tests, but the method of discussing the concept questions will be different for the two different groups.

Number of Participants

About 100 people will take part in this study.

Procedures

If you agree to allow your child to be in this study, then you or your child will not have any additional requirements. By agreeing to allow your child to participate you are agreeing to allow your child's scores on the Force Concept Inventory and the benchmark tests to be reported in the study once the scores have been assigned to a random number to protect the identity of your child, the participant.

Length of Participation

The study will be conducted during eleven weeks of the Fall 2011-2012 school year. The assignment to the groups is based on which hour your child has the class.

This study has the following risks:

Since both groups are covering the same material there should be no additional risk for your child.

Benefits of being in the study are

that students in both groups will spend time discussing the major concepts from the class during the normal class period sparked by Eric Mazur's conceptual questions from his Harvard physics class.

Confidentiality

In published reports, there will be no information included that will make it possible to identify your child without your permission and your child's permission. Research records will be stored securely and only approved researchers will have access to the records.

There are organizations that may inspect and/or copy your research records for quality assurance and data analysis. These organizations include the Instructional Leadership and Academic Curriculum Department and the OU Institutional Review Board.

Compensation

Your child will not be reimbursed for his or her time and participation in this study.

Voluntary Nature of the Study

Participation in this study is voluntary. If you withdraw or decline your child's participation, your child will not be penalized or lose benefits or services unrelated to the study. If you decide to allow your child to participate, your child may decline to answer any question and may choose to withdraw at any time.

Request for record information

If you approve, your child's confidential records will be used as data for this study. The records that will be used include GPA, gender, age, and math classes taken. These records will be used for the following purpose(s): All records will be used to correlate or see how much certain variables account for the results by grouping students with similar GPA, math classes, or other variables.

_____ I agree for my child's school records to be accessed and used for the purposes described above.

_____ I do not agree for my child's school records to be accessed for use as research data.

Contacts and Questions

If you have concerns or complaints about the research, then Mr. Warren can be contacted through email at [redacted] or by phone at [redacted]. Timothy Laubach the faculty advisor from OU can also be contacted at [redacted] or by phone at [redacted].

Contact the researcher(s) if you have questions or if your child has experienced a research-related injury.

If you have any questions about your child's rights as a research participant, concerns, or complaints about the research and wish to talk to someone other than individuals on the research team or if you cannot reach the research team, you may contact the University of Oklahoma – Norman Campus Institutional Review Board (OU-NC IRB) at 405-325-8110 or irb@ou.edu.

You will be given a copy of this information to keep for your records. If you are not given a copy of this consent form, please request one.

Statement of Consent

I have read the above information. I have asked questions and have received satisfactory answers. I consent to allow my child to participate in the study.

Student Name

Parent Signature Date

University of Oklahoma
Institutional Review Board
Student Assent to Participate in a Research Study

Project Title: Concept Questions with Peer-led vs. Teacher-led Discussions in a High School Physics Classroom
Principal Investigator: Kevin Warren
Department: Instructional Leadership and Academic Curriculum

You are being asked to volunteer for this research study. This study is being conducted at [redacted] High School. You were selected as a possible participant because you are in Mr. Warren's physics class this semester. Please read this form and ask any questions that you may have before agreeing to take part in this study.

Purpose of the Research Study

The purpose of this study is to compare a class with peer-led discussions versus a class with teacher-led discussions. The study will compare the learning gains for students in each group on the Force Concept Inventory and the benchmark tests. Both classes will have the same labs, worksheets, and tests, but the method of discussing the concept questions will be different for the two different groups.

Number of Participants

About 100 people will take part in this study.

Procedures

If you agree to be in this study, then you will not have any additional requirements. By agreeing to participate you are agreeing to allow your scores on the Force Concept Inventory and the benchmark tests to be reported in the study once the scores have been assigned to a random number to protect your identity.

Length of Participation

The study will be conducted during eleven weeks of the Fall 2011-2012 school year. The assignment to the groups is based on the hour of your class.

This study has the following risks:

Since both groups are covering the same material there should be no additional risk.

Benefits of being in the study are

that students in both groups will spend time discussing the major concepts from the class during the normal class period sparked by Eric Mazur's conceptual questions from his Harvard physics class.

Confidentiality

In published reports, there will be no information included that will make it possible to identify you without your permission. Research records will be stored securely and only approved researchers will have access to the records.

There are organizations that may inspect and/or copy your research records for quality assurance and data analysis. These organizations include the Instructional Leadership and Academic Curriculum Department and the OU Institutional Review Board.

Compensation

You will not be reimbursed for your time and participation in this study.

Voluntary Nature of the Study

Participation in this study is voluntary. If you withdraw or decline participation, you will not be penalized or lose benefits or services unrelated to the study. If you decide to participate, you may decline to answer any question and may choose to withdraw at any time.

Request for record information

If you approve, your confidential records will be used as data for this study. The records that will be used include GPA, gender, age, and math classes taken. These records will be used for the following purpose(s): All records will be used to correlate or see how much certain variables account for the results by grouping students with similar GPA, math classes, or other variables.

_____ I agree for my school records to be accessed and used for the purposes described above.

_____ I do not agree for my school records to be accessed for use as research data.

Contacts and Questions

If you have concerns or complaints about the research Mr. Warren can be contacted through email at [redacted] or by phone at [redacted]. Timothy Laubach the faculty advisor from OU can also be contacted at [redacted] or by phone at [redacted] [redacted].

Contact the researcher(s) if you have questions or if you have experienced a research-related injury.

If you have any questions about your rights as a research participant, concerns, or complaints about the research and wish to talk to someone other than individuals on the research team or if you cannot reach the research team, you may contact the University of Oklahoma – Norman Campus Institutional Review Board (OU-NC IRB) at 405-325-8110 or irb@ou.edu.

You will be given a copy of this information to keep for your records. If you are not given a copy of this assent form, please request one.

I have explained the study to ______________________(*print name of child here*) in language he/she can understand, and the student has agreed to be in the study.

______________________________ ________________

Signature of Student Date

______________________________ ________________

Signature of Person Conducting Assent Discussion Date

____

Name of Person Conducting Assent Discussion (*print*)

This volume is the property of the University of Oklahoma, but the literary rights of the author are a separate property and must be respected. Passages must not be copied or closely paraphrased without the previous written consent of the author. If the reader obtains any assistance from this volume, he must give proper credit in his own work.

I grant the University of Oklahoma Libraries permission to make a copy of my thesis upon the request of individuals or libraries. This permission is granted with the understanding that a copy will be provided for research purposes only, and that requestors will be informed of these restrictions.

NAME [redacted]
DATE

A library which borrows this thesis for use by its patrons is expected to secure the signature of each user.

This thesis by KEVIN WARREN has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

NAME AND ADDRESS	DATE