INTERACTIVE-ENGAGEMENT VS. COOKBOOK LABORATORY PROCEDURES IN MBL MECHANICS EXERCISES

by

Brent Royuk

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Interdepartmental Area of Administration, Curriculum & Instruction
Under the Supervision of Professor David W. Brooks

Lincoln, Nebraska
February, 2002
INTERACTIVE-ENGAGEMENT VS. COOKBOOK LABORATORY PROCEDURES IN MBL MECHANICS EXERCISES

Brent Royuk, Ph.D.
University of Nebraska, 2002

Advisor: David W. Brooks

In recent years, much work has been done to investigate physics teaching techniques that facilitate conceptual learning in mechanics. This study compared the effectiveness of microcomputer-based laboratory procedures that were written in a traditional “cookbook” style to interactive-engagement procedures that covered the same material with the same experimental apparatus for equal times.

Two lab sections in an introductory trig-based physics course at a small private college participated in different lab exercises for nine weeks. One section completed nine chapters of the interactive-engagement lab curriculum, RealTime Physics. The other participated in cookbook labs that were written for this study to cover the same material. Gain in conceptual mechanics understanding was measured with a pre-instruction/post-instruction administration of the Force Concept Inventory. Both groups completed the conceptual homework included in the RealTime Physics exercises. This procedure was repeated in a second nine-week phase, in which neither group was assigned the homework.

Average normalized gains for the interactive-engagement and cookbook groups were $h = 0.471$ and $h = 0.392$, respectively. In the second phase (without the homework), they were $h = 0.480$ and $h = 0.334$. In the second phase, the normalized
gain for the interactive-engagement group was 0.568 s.d. higher than the cookbook
group \((N = 27, p = 0.076)\).

For the interactive-engagement groups in the two phases, the homework did not
make a difference in FCI gains. The pooled average normalized gain for these two
groups was equal to \(h = 0.476 (N = 27)\), which is comparable to the average gain
measured for the interactive-engagement groups in Hake’s large data set in 1998.

Small differences in satisfaction and perceived effectiveness were measured
between the interactive-engagement and cookbook groups. These differences generally
favored the cookbook labs.
ACKNOWLEDGEMENTS

I owe a debt of thanks to many people who helped during the planning, implementation and analysis phases of this study. I would first like to thank the members of my doctoral committee: Charles Ansorge, David Fowler, and Diandra Leslie-Pelecky. Thanks to each of you for your helpful advice and wise counsel.

Thanks to Tom Koch and Vicki Plano Clark for serving on the lab-writing advisory panel. You waded through a deep pile of .pdf files while we were writing the cookbook labs and gave me expert advice that made the project much better.

I’m very grateful to Rob Hermann, who acted as the instructor for the test courses and graciously gave up control of his laboratories in order to do this project. Thanks also to our able team of lab assistants, Jon Sprengel, Abby Duthie and Josh Dixon. I’d also like to thank Kay Rohren, Valarie Eilert and Rebecca Jabs for secretarial help during the study.

This work was also supported by a grant from the Nebraska Catalyst Project. Thank you for making my research possible.

My greatest thanks are reserved for my mentor David Brooks, who pushed, prodded, and pulled me all the way through my doctoral and dissertation-writing process. You’ve had a big effect on me, Dave, and that’s what teaching is all about.

Finally, I wish to thank my dear wife, Sandra, who has been my confidante, supporter and soul-mate through ten years of graduate school. Raising four children during the same period of time isn’t the easiest way to do things, but it’s been worth it.

Soli Deo Gloria
# TABLE OF CONTENTS

## I. INTRODUCTION

- Context 1
- Purpose of the Study 2
- Research Questions 3
- Significance of the Study 4

## II. REVIEW OF LITERATURE

- Introduction 5
- Historical Perspectives on Educational Reform 5
- Overview of the Current State of Physics Education Research 7
  - Constructivism 7
  - Resistance to Reform 9
  - Breadth vs. Depth 9
  - Filter or Pump? 10
- The Reform Agenda 11
- The Force Concept Inventory 12
  - Origins of the FCI 13
- Assessing Instructional Reform 13
  - Richard Hake’s Study 14
  - Other Uses of the FCI 16
- Validity and Reliability of the FCI 17
  - Validity 18
  - Reliability 18
  - Huffman and Heller’s Argument 19
  - Summary 19
- Educational Research on the Science Laboratory 20
- Physics Education Research Laboratory Perspectives 20
  - Thornton and Sokoloff’s First Paper 21
  - Redish, Saul, and Steinberg’s Paper 21
  - RQ1 in the Literature 22
  - Thornton and Sokoloff’s Second Paper 23
  - Other Physics Laboratory Studies 23
- Cookbook Labs and Inquiry Learning 24
  - Inquiry in the National Standards 25
  - Definition of Interactive-Engagement Laboratories 26
- Cookbook Labs Defined 26
  - Definition of Cookbook Laboratories 27
- MBL in Hake’s Study 28
  - Hake’s Case Studies and RQ2 29
- Student Satisfaction Surveys 29
III. METHODS

Population and Sample
Variables and Measures
Limitations
Procedural Steps
Treatments
  Weekly Schedule
  Selection of Chapters
  Conduct of the Laboratory Activities
The Student Satisfaction Survey
Instructor and Lab Assistants
Writing the Cookbook Labs

IV. RESULTS

The Hake Factor
Average Normalized Gains
T-Test Results
Student Satisfaction Survey

V. DISCUSSION

Research Question #1
  Are the Cookbook Labs Realistic?
  The Instructor’s Critique
  “Traditional” vs. “Cookbook”
  Length of Treatment
Research Question #2
Research Question #3
Suggestions for Further Research

VI. Summary

REFERENCES

Appendix A  The Student Satisfaction Survey
Appendix B  Cookbook Laboratory Exercises
Appendix C  Review Panel Instructions
Appendix D  Instructor’s Feedback
Appendix E  IRB Documents
<table>
<thead>
<tr>
<th>Figure/Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>A schematic diagram of the Hake Plot.</td>
<td>15</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Pie charts that summarize the year in school and majors for participants in the study (N = 52).</td>
<td>31</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Exercises in RealTime Physics: Mechanics completed by the interactive-engagement laboratory section.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Hake Plot for individuals (N = 52) in each treatment group. Phase I included homework, Phase II did not.</td>
<td>41</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Summary of pretest scores, posttest scores, and mean normalized gains for the four experimental groups.</td>
<td>42</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Bar graph of mean normalized gains for the four experimental groups.</td>
<td>42</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Hake plot for group averages. The two dashed lines divide the allowable region into areas of high-gain, medium-gain, and low-gain. The two thin solid lines show the average normalized gains for the 14 traditional courses and the 48 IE courses in Hake’s study, which are included in the plot.</td>
<td>43</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Results of unpaired t-tests between the IE and cookbook groups in both phases of the experiment.</td>
<td>44</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Abbreviations used for reporting satisfaction survey results.</td>
<td>45</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Average response scores to the Student Satisfaction Survey for the four experimental groups. High scores correspond to positive responses.</td>
<td>45</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Bar graph of satisfaction survey responses by treatment group.</td>
<td>46</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Bar graph of satisfaction survey responses by type of treatment group.</td>
<td>47</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Bar graph of satisfaction survey responses by term, i.e. whether or not the participants also did the homework.</td>
<td>48</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Summary of student responses to the first free-response question on the opinion survey.</td>
<td>49</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Summary of student responses to the second free-response question on the opinion survey.</td>
<td>49</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>Summary of student responses to the third free-response question on the opinion survey.</td>
<td>50</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Context

The educational laboratory has been used as an instructional tool in the physics classroom for many years. As early as 1886, Harvard University published a list of physics experiments to be completed by high school students who wished to enroll at Harvard (Moyer, 1976). Physics teachers recognize the importance of educational laboratory exercises in assisting the acquisition of laboratory skills, introducing the processes of scientific inquiry, and as an instructional strategy to help students learn physics concepts.

There have been many physics curricula developed for introductory courses in recent years. Well-known current curriculum projects making use of specialized laboratory resources include RealTime Physics (Sokoloff, Thornton, & Laws, 1999), Tools for Scientific Thinking (Thornton, 1987), Socratic Dialogue Inducing Labs (Hake, 1987), and The Modeling Workshop Project (Halloun & Hestenes, 1987). Many of these modern laboratory curricula have been developed using results from physics education research. The laboratory procedures in these exercises differ in many ways from traditional labs. In addition to the increased availability of technological tools in the modern labs, these laboratory exercises require the learner to be an active inquirer, solving problems and focusing attention on the ideas and processes of physics.

Old-style labs, on the other hand, are often derided in the research literature as “cookbook” labs. The lab manuals of twenty years ago generally contained explicit
instructions in a step-by-step procedure that choreographed each action taken by the learner, with reflective questions saved for the end of the lab exercise. After “taking data,” the students were often required to write a report in some prescribed format, where they would be expected to synthesize the main points of the exercise into some sort of cohesive conceptual aggregate.

Physics laboratory curricula written in the last decade often include the use of Microcomputer Based Laboratory (MBL) equipment. These curricula were originally developed by educators at postsecondary schools, especially Tufts University (Thornton & Sokoloff, 1990). Measuring devices that interface with a computer give students in the physics laboratory unprecedented access to real-time measurements of physical phenomena.

Purpose of the Study

Physics education researchers have made progress in acquiring an understanding of instructional strategies that improve the conceptual learning of physics students. But much of the data that has been collected has failed to isolate single instructional treatments from instructional regimes that use a variety of teaching strategies. This study has focused on one particular type of instructional tool in physics, the educational laboratory. By isolating this treatment in a controlled experiment, an attempt has been made to determine its importance in producing a change in conceptual mechanics knowledge.

Several researchers have considered laboratory exercises as a separate factor in student learning. Some, for instance, have compared the effectiveness of modern MBL labs to traditional lecture instruction (Redish, Saul, & Steinberg, 1997). Some data also
supports the effectiveness of the use of MBL instructional methods for conceptual learning gains (Workshop Physics project, 2001).

Perhaps the effectiveness of these laboratory curricula resides in their use of MBL equipment itself. Or is the enhanced effectiveness due to the fact that learners are actively engaged by the procedures in this sort of laboratory? Would physics students learn mechanical concepts just as efficiently in a cookbook lab that makes use of MBL instruments? Does the benefit reside in the active-learning characteristics of the procedure? This study has tried to separate these issues.

Research Questions

In this research study, the following questions were investigated:

RQ1. Are there significant differences in the conceptual mechanics knowledge gain (as measured by the FCI) for students who participate in active-learning MBL physics laboratories, compared to students who participate in equal-time exercises with cookbook procedures that also make use of MBL equipment?

RQ2. Can the use of interactive-engagement laboratories in conjunction with an otherwise traditional classroom environment produce significant gains in conceptual learning?

RQ3. How do the satisfaction and perceived effectiveness of the exercises compare for students in the two groups?
Significance of the Study

While many modern active-learning instructional techniques have been widely adopted, there is still much resistance to change. There are few, if any, active proponents of cookbook labs in the physics education research community. However, it is likely that many individual physics instructors continue to use traditional lab exercises that rely on cookbook procedures. It is therefore important to determine if there is an educational benefit to using active-learning procedures in the educational physics laboratory that surpasses the benefit produced by these traditional procedures.

Many physics teachers who cling to traditional instructional techniques use some kind of laboratory exercises, though they may possess characteristics that would be described as “traditional.” Previous studies that have compared active-learning MBL labs to lecture have not made a comparison with traditional labs. The treatment factor varied in this experiment was the engagement level of the laboratory procedure. By isolating this particular characteristic, this study provides a direct comparison of a course that makes use of MBL equipment and interactive-engagement techniques in its laboratory component, and one that uses the newest equipment with old-fashioned instructional procedures.

In summary, this study will evaluate the effectiveness of active learning procedures in the introductory physics laboratory while holding constant as many other factors as possible. This will afford individual physics instructors more information upon which to make decisions about the nature of the laboratory exercises they provide for their students.
CHAPTER II

REVIEW OF LITERATURE

Introduction

Science educators are constantly striving to improve the quality of education. In this chapter, historical reform efforts will be described and current ideas about instructional theory will be summarized. In order to localize the present study in a research base, this review of literature will focus on educational research efforts in the field of physics, and will particularly emphasize Newtonian mechanics.

In the field of mechanics, physics education researchers have a unique advantage over their counterparts in many other disciplines, in that there exists a short, widely used evaluation instrument, namely the Force Concept Inventory. This chapter will review the development and history of the Force Concept Inventory, synopsize recent research in conceptual mechanics education, and finally discuss research studies that have focussed on the educational physics laboratory.

Historical Perspectives on Educational Reform

Trowbridge, Bybee, & Powell identify two distinct eras of reform in the last forty years. A “Golden Age” took place roughly from 1958-1988, and a “Modern Era” dates from 1988 to the present (2000). Each of these eras has been characterized by a particular psychological theory that has guided its development. The Golden Era featured an emphasis on behavioral psychology and Piaget’s theory of cognitive development, while the Modern Era stresses constructivism and inquiry learning.
The Golden Era (1958-1988). The Golden Era of reform was ushered in by the launching of Sputnik in 1957, which caused a huge influx of federal money for the development of new science curricula and a large increase in national attention to the importance of science education. These reforms followed an earlier round of reform that was precipitated by World War II (Donahue, 1993). During the early stages of the Cold War, educators were pushed by federal policy to produce unambiguously measurable learning outcomes. In this era science educators moved away from the curriculum strategies of the first half of the century that accentuated technical and social applications of technology, and moved toward learning the abstractions and theories of science.

During the 1960s, an alphabet soup of curriculum programs was produced. These included the Physical Science Study Committee (Physical Science Study Committee, 1957), the Earth Science Curriculum Project (Earth Science Curriculum Project, 1965), and several others which facilitated students’ acquisition of scientific knowledge.

A typical study during the 1970s (Griffiths, 1975) measured the cognitive development level of students studying physics, and studied whether a level of development had been attained that would allow understanding of physics concepts. The author concluded that many students had not reached Piaget’s formal operation level of cognitive development and could therefore actually be “harmed” by physics instruction.

Prigo (1978) attempted to develop a lecture course that was sensitive to the cognitive development level of its students. Given that about 50% of college freshmen still operate at the concrete operational level of thinking, he felt it was necessary to focus the attention of his course on the “objects” of physics, before it was appropriate for the theories of the discipline to be considered.
Liberman & Hudson (1979) measured a correlation \( r = 0.49 \) between logical abilities and academic achievement in physics. They suggested that formal operation reasoning abilities are a necessary precondition to learning physics.

Classifying students by cognitive ability level and allowing the “cream to rise to the top” does not mesh with prevailing notions of instruction. Present researchers focus on the methods of instruction, searching for techniques that will be helpful to students at all developmental levels.

**Overview of the Current State of Physics Education Research**

The state of the current reform movement in physics education has been summarized by Mestre (1994). The main obstacle to learning physics he describes in his paper is highly representative of that found in much of the present literature, namely student misconceptions.

**Constructivism.** In constructivist theory– the dominant paradigm among science education researchers– all knowledge needs to be “constructed” by students in a highly contextual way. Students do not come to science classes with a “blank slate,” ready to have knowledge transmitted to them by a content expert, but rather need to relate new knowledge to that which is already present in their minds. This process is hampered by the presence of misconceptions, or naïve beliefs. These beliefs need to be “flushed out” into the open by some socially mediated process; a dialogue with the teacher, small group discussions, experience with manipulable items in the laboratory, or a combination of these techniques. Once the students are confronted with the inadequacy of their misconceptions, they can begin the process of building accurate conceptual knowledge.
Constructivism is a movement with many diverse proponents, so it is not surprising that many differing opinions exist as to what constructivism is and what it is not. Brooks and Brooks (1993) offer one perspective, by listing “Five Principles of Constructivist Classrooms,” as:

1. Teachers seek and value their students’ point of view.
2. Classroom activities challenge students’ suppositions.
3. Teachers pose problems of emerging relevance.
4. Teachers build lessons around primary concepts and “big” ideas.

This learning theory is compatible with the “inquiry learning” instructional approach, which focuses on helping students pose answerable questions, devise a procedure to answer the question, and communicate the results (Trowbridge et al., 2000). Various national science standards documents endorse inquiry learning as an effective and important instructional strategy (American Association for the Advancement of Science, 1989; National Research Council, 1996).

Much effort has been devoted to the discovery and measurement of misconceptions in the general population. Misconceptions have been found to be very common, even among science teachers. It is important that instructors are aware of the nature of these misconceptions, so that they can be properly addressed during instruction.

This approach contrasts with traditional instruction, which takes a “transmittalist” approach to instruction. In this system, information is presented to the students through lectures, while students sit passively and absorb it. Transmittalists assume that success in
learning largely depends on the clarity of the presentation, and the charisma of the teacher (McDermott, 1999).

**Resistance to Reform.** Reformers often wrestle with the fact that their agenda remains uncommon in modern science classrooms. Redish (2000) states, for example, that “Although there has been an intellectual explosion in physics curriculum development, the actual impact on teaching at the tertiary level has so far been small. Most innovations remain local, ignore the results of physics education research and cognitive science, and are ineffective.” Mestre (1994) offers two reasons that account for the same problem: teachers are merely continuing a “vicious cycle,” teaching as they were taught, and that they are overwhelmed by the need to cover an ever-increasing amount of material. Redish, on the other hand, speculates that the main obstacle to the implementation of reform comes from basic misconceptions that instructors have about how students learn.

**Breadth vs. Depth.** Mestre’s second point underscores another fault line between the modern reform era and the previous reform era: breadth versus depth. In the new era of science education reform, “less is more” (Speece, 1993). Many reformers argue that it is more important to cover the meaty concepts of a discipline in depth, to avoid a curriculum that is “a mile wide and an inch deep.” (Schmidt, McKnight, & Raizen, 1997) Some have characterized this debate as “the religious question of whether it’s better to learn 10% of 90% of the subject or 90% of 10% of the subject.” (Brooks, 2000) Brooks and Brooks speak to this issue very clearly, by saying

**Constructivist teachers have discovered that the prescribed scope, sequence, and timeline often interferes with their ability to help students understand complex**
concepts. Rigid timelines are also at odds with research on how human beings form meaningful theories about the ways the world works (Duckworth 1986), how students and teachers develop an appreciation of knowledge and understanding (Eisner 1985), and how one creates the disposition to inquire about phenomena not fully understood (Katz 1985). Most curriculums simply pack too much information into too little time—at a significant cost to the learner. Teachers everywhere lament how quickly students forget and how little of what they initially remembered they retain over time. Our present curricular structure has engineered that outcome. Students haven't forgotten; they never learned that which we assumed they had. In demanding coverage of a broad landscape of material, we often win the battle but lose the war. We expose students to the material and prepare them for the tests, but we don't allow them to learn the concepts. (p. 39-40)

Filter or Pump? Although the philosophical underpinnings of the two reform eras were instrumental in producing the differences described above, one could argue that the driving force that has produced the shift to the modern paradigm was not theoretical, but rather a change in goals. The Sputnik-inspired reforms of the 1960s were designed to focus primarily on the high-achieving students bound for college, in order to produce an elite cadre of scientists and engineers who would ensure the technological superiority of the United States for years to come. An extreme example of this point of view was voiced in 1942:

Excoriating the “extreme phase of mass and moron worship” which he saw in the public schools, Thomas Cope of the University of Pennsylvania believed physics
should be used to separate the elite from “the horde of less intelligent pupils which today overcrowds the public secondary schools.” He asked, “Is our high school boy able to master Millikan and Gale’s *Physics* and is he willing to make the necessary effort? If yes he is my aristocrat, if not, he belongs to my masses.” (Cope, 1942; quoted in Donahue, 1993, p. 330)

When voices are raised in opposition to current reform initiatives, the point of disagreement often breaks down to this question of goals: Do we wish to continue broad traditional coverage of topics, which may act as a “filter,” weeding out low-achievers, or a “pump” (Steen, 1988), following the dictates of the National Science Education Standards (National Research Council, 1996), which call for “science for all?” Anti-reformers might object to the present emphases by saying “What’s wrong with the old methods? I learned fine that way,” to which a reformer would reply “But that’s how you learned it. Were you an average student? What about the medium and low-achievers who need to learn science?” The debate can be contentious, with one author sarcastically asking “Why Change, Been Doin’ It This Way For 4000 Years!” (Flowers, 2000)

**The Reform Agenda**

Researchers propose several strategies to help students acquire the skills and strategies necessary for learning physics. First of all, teachers need to be aware of the obstacles students face when learning physics. This implies they need focus on the learner, and not only display competence in their subject. For instance, if an instructor doesn’t realize that some students lack prerequisite math skills, those students are likely to fail. The teacher therefore needs to solicit feedback from the students, in order to help them overcome any skill deficiencies they might have. This “feedback principle” is also
of prime importance in helping students overcome their misconceptions. Merely distributing information without engaging students in any kind of active learning activities that address misconceptions will increase the likelihood of failure.

Teaching methods seem to be of extreme importance in helping students acquire conceptual knowledge about mechanics. In important papers that will be considered in some detail below (Hake, 1998a; Hake, 1998b), Hake has shown instructional methods lacking elements of “interactive engagement” to be ineffective in helping students acquire conceptual knowledge, as measured by instruments like the Force Concept Inventory. These results have led Hestenes (1998) to suggest that “lectures are (perhaps, totally) ineffective in teaching the basic concepts of physics, even apart from other evidence pointing to the same conclusion.” (p. 466) Similarly, Redish states that “as physics teachers we fail to make an impact on the way a majority of our students think about the world.” (1994, p. 796).

The Force Concept Inventory

The Force Concept Inventory (FCI) is a unique instrument. In a post to the Classics-L listserv on May 25, 1999, Tompkins wrote:

The neat thing about physics is that there is a pretty good instrument called the Force Concept Inventory, which basically tests student understanding of principles that are counter-intuitive-- that is, it is a measure not of “binge-and-purge” learning but of deeper understanding.

The above quote demonstrates that the FCI has even become known outside the physics education community. Within that community it is the most widely used diagnostic tool in existence, and has been cited in dozens of research articles.
**Origins of the FCI.** The FCI was introduced by Hestenes, Wells and Swackhamer in 1992. This instrument evolved from the earlier Mechanics Diagnostic Test (Halloun & Hestenes, 1985a) and was revised slightly in 1995. Consisting of thirty multiple-choice questions, the FCI was designed to measure students’ conceptual mastery of Newtonian mechanics. Although experienced physics problem-solvers familiar with the FCI generally find answers to the questions to be obvious and indisputable, novices score very poorly on the instrument, particularly since the distracter responses were designed to match common misconceptions.

The FCI is commonly used to measure the effect of an educational treatment, through pre-instruction and post-instruction administration. In this type of design, educational researchers measure the pretest and posttest scores in order to calculate the gain achieved through some instructional treatment regime.

By 1998, David Hestenes had data from more than 20,000 students in 300 physics classes, ranging from high school to graduate school (Hestenes, 1998). Eric Mazur has included the FCI in his popular book, *Peer Instruction* (1997). The instrument is unique in its ubiquity. The Conceptual Astronomy and Physics Education Research Team at Montana State University even advertises its Astronomy Diagnostics Test as the “FCI for astronomy” on its departmental webpage (Montana State University, 2001).

**Assessing Instructional Reform**

The FCI has been an important tool in assessing the effectiveness of various educational treatments in introductory-level physics courses. Halloun and Hestenes (1985a) summarized the findings of several studies from the early 1980s regarding common-sense beliefs about motion as:
1. Common sense beliefs about motion are generally incompatible with Newtonian theory. Consequently, there is a tendency for students to systematically misinterpret material in introductory physics courses.

2. Common sense beliefs are very stable, and conventional physics instruction does little to change them. (p. 1043)

Richard Hake’s Study. The search for the most effective means to bring about changes in these beliefs eventually led Richard Hake (1998a; 1998b) to compare gains in the FCI and Mechanics Diagnostic Test for courses that used “traditional” instruction to those using “interactive-engagement” methods.

Hake solicited data from sixty-two introductory physics courses enrolling a total of 6542 students. Forty-eight of the courses were classified as interactive-engagement courses, with the remaining fourteen labeled as traditional. Hake averaged the pretest and posttest scores for each course in the two groups and used this data to calculate the average normalized gain for each class, $\text{Gain}_{\text{normalized}} = h = \frac{\text{post} - \text{pre}}{1 - \text{pre}}$ (where $\text{pre}$ and $\text{post}$ are the ratios of the number correct to the total possible), which he then plotted vs. the pretest score. This diagram has come to be known as a Hake Plot, and the normalized gain is often referred to as the Hake Factor, $h$ (Francis, Adams, & Noonan, 1998; Redish et al., 1997). A schematic of the Hake Plot is shown in Figure 2.1. Since the gain of each student is limited to 100% minus the percentage earned on the pretest, all points lie in the shaded triangular region. Hake showed that classes with similar instructional approaches tend to lie on straight lines passing through the point (100,0). High-gain courses that make use of interactive-engagement instructional methods lie closer to the
top of the shaded region (Line a in the diagram), while traditional instruction tends to produce gains that lie on lines closer to the horizontal axis (Line b).

![Hake Plot Diagram](image)

Figure 2.1. A schematic diagram of the Hake Plot. Class average gains are plotted against pretest scores. Interactive-engagement courses tend to cluster along lines closer to the top of the shaded region, with traditional courses closer to the bottom.

In his paper, Hake defined interactive-engagement methods as “*those designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors*” (p. 65). Hake includes MBL exercises as one of his interactive engagement components. The interactive engagement group of courses experienced an average gain of $0.48 \pm 0.14$ (std. dev.), while the traditional courses had an average gain of $0.23 \pm 0.04$ (std. dev.). The difference in these
average gains, $0.48 - 0.23 = 0.25$, represents an effect size of 1.8 standard deviations of the interactive-engagement courses (0.14) and 6.2 standard deviations of the traditional group (0.04).

No challenges to Hake’s study have appeared in the literature. A significant amount of discourse regarding the validity of the Hake study is found on electronic bulletin boards. For example, in a post to the PHYS-LRNR listserv Ron Greene stated that “The statement below [that Hake’s study demonstrated substantial learning gains with interactive-engagement methods] has not YET been demonstrated because active vs. inactive learning was not the only relevant distinction between the two groups.” (2000) Hake responded to Greene’s statement in a subsequent listserv post, retracting the word “demonstrated,” but defending the methodology of the study (Hake, 2000). Issues in Hake’s study that are concerned with laboratory work in particular will be considered in more detail in a following section.

Other Uses of the FCI. Francis, Adams & Noonan (1998) administered the FCI to students who had taken an introductory physics course as many as four years earlier, in order to see if high scores on the exam remained fixed. Since there was such a long delay following instruction, the authors reasoned that students who had scored well by memorizing the “right answers” would have been likely to forget them by then. If, on the other hand, high scores had been achieved through the acquisition of a truly Newtonian worldview, the scores would remain high. With $N = 127$, they found that FCI scores were lower by an average of only seven percentage points, with the students having the greatest delay since taking the course actually showing the smallest difference. The
authors concluded that this persistence of scores suggests an enduring shift in beliefs about motion.

Saperstein (1995) attempted to determine if FCI scores tend to rise for early teenage students through “living,” without receiving any sort of formal physics instruction. By comparing FCI scores for a group of 40 girls of age 12 ± 0.5 to published pretest scores of high school seniors and college freshmen, Saperstein concluded that students experience a gain in FCI scores of (2.3 ± 1.2)% merely through everyday life.

Other researchers have considered the format of the FCI in particular. These include two studies that have compared multiple-choice responses to free-response answers on the FCI (Rebello & Zollman, 2000; Steinberg & Sabella, 1997), and one that used animated applet simulations on the exam’s answers, in place of static printed alternatives (Dancy, Titus, & Beichner, 2000).

**Validity and Reliability of the FCI**

When Hestenes, Wells & Swackhamer developed the FCI, they identified six conceptual dimensions as part of a comprehensive Newtonian force concept (1992):

1. Kinematics
2. First Law
3. Second Law
4. Third Law
5. Superposition Principle
6. Kinds of Force (p. 142)

Items in the exam are keyed specifically to each of these six dimensions. In each dimension, commonsense misconceptions, as collected in student interviews by Halloun
and Hestenes (1985b), were used to generate distracter options in the corresponding items. There are six categories in which commonsense misconceptions occur, those being:

1. Kinematics
2. Impetus
3. Active Force
4. Action/Reaction Pairs
5. Concatenation of Influences
6. Other Influences on Motion (Hestenes et al., 1992, p. 143-145)

Validity. To validate the FCI, the authors made use of validation work done on its precursor, the Mechanics Diagnostic Test (Halloun & Hestenes, 1985a). In this process, a draft of the test was shared with physics professors and graduate students, adopting some of their suggested revisions. The test was then given to a panel of graduate students, all of whom were able to agree on the correct answers. Interviews were then conducted with high school students who had taken the test, to test for understanding of the questions. Then tests taken by “A” students in a University Physics course were examined for patterns of common misunderstandings.

Reliability. The reliability of the test was established through interviews with students who had taken the test, with the investigators finding excellent agreement between the way students thought and the answers they gave. The Modeling Workshop project at Arizona State University has collected FCI data from some 20,000 high school students and has calculated Cronbach’s coefficient alpha, which measures the instrument’s reliability. They have obtained coefficient alpha values of “mid .80s to the
mid .90s” for FCI posttests and “high .60s to mid .70s” for FCI pretest scores. (Popp, 2000). The alpha coefficient varies from 0 to 1, where higher values indicate greater reliability, and values in the range of 0.7 to 0.8 are generally considered acceptable in social science research (Santos, 1999).

**Huffman and Heller’s Argument.** Huffman and Heller (1995) used a factor-analysis technique to argue that there is not actually a single “force concept” that is measured by the FCI. They examined FCI test data for 145 high school and 750 university students and found that there were only three clusters of questions that grouped together in a statistically significant way. These clusters were not strongly associated with single conceptual dimensions in the FCI. From this they concluded that “the items on the FCI are only loosely related,” (p. 140). They further assert that since students have ill-formed Newtonian mechanical concepts, the FCI is unable to measure a unified concept, but only measures “bits and pieces” of student understanding.

Responding to this, Hestenes and Halloun (1995) said that this is to be expected, since the data was collected from a non-Newtonian population, and the subjects therefore have no Newtonian force concept that can be measured.

The argument continued through two other papers (Halloun & Hestenes, 1996; Heller & Huffman, 1995). Much common ground was established in these subsequent papers, but neither group conceded the main points of the argument.

**Summary.** Given its reliability, validity and popularity, the FCI was chosen as the assessment tool for this project. There is also a significant amount of published data that uses the same test, so results from this study can easily be compared to previous research.
Educational Research on the Science Laboratory

Blosser (1983) surveyed research dealing with the role of the laboratory in science education. She states that the educational laboratory has been a common feature of introductory courses since the 1800s, and has received special emphasis during the reforms of the 1960s. According to Blosser, teaching laboratories are used to attain a wide variety of objectives, beyond merely acquiring content. These include attitudinal goals, familiarity with tools and techniques, and adding reality to the material in the textbook. The educational laboratory also has had its detractors, who feel that laboratory exercises may not present a clear picture of how real science is conducted (Blosser, 1983).

In Blosser’s opinion, much of the literature dealing with educational laboratories express opinions rather than research-based facts. She feels that too many of the research studies are doctoral dissertations that represent an individual’s first attempt at research, and do not include any follow-up studies. Many of the studies failed to detect statistically significant differences between educational treatments.

On a positive note, Blosser cites several studies that clearly demonstrate the effectiveness of laboratory activities, including a study by Comber and Keeves (1978) that compared science education in 19 countries, and found higher achievement levels in countries that made use of teaching laboratories.

Physics Education Research Laboratory Perspectives

In the context of the modern physics education research framework, several papers have investigated the effect of laboratory work on student learning.
Thornton and Sokoloff’s First Paper. Thornton & Sokoloff described a kinematics curriculum that was implemented with early MBL devices (Thornton & Sokoloff, 1990). Making use of a pretest/posttest design, they demonstrated the effectiveness of MBL-aided laboratory exercises. They concluded that MBL tools by themselves did not necessarily produce conceptual understanding, but that “These gains in learning physics concepts appear to be produced by the combination of the tools and the appropriate curricular materials.” (p. 865)

Redish, Saul, and Steinberg’s Paper. A study at the University of Maryland (Redish et al., 1997) attempted to extend the work of Thornton and Sokoloff by using MBL activities, controlling the time spent on the topic and probing the problem-solving ability of the students in the study. Engineering students in an introductory physics course were divided into two groups. One group of five lecture classes participated in recitations while the other group of five lecture classes participated in two MBL “tutorials” dealing with the concept of instantaneous velocity and Newton’s third law. Students were evaluated using the multiple-choice velocity questions developed by Thornton and Sokoloff, the FCI, and one long-answer question.

Although the treatment group only participated in tutorials dealing with instantaneous velocity and Newton’s third law, the whole FCI was administered in a pretest/posttest format to “provide a normalization of the overall effectiveness of the tutorial environment for general concept building.” (p. 48) It was found that the tutorial classes experienced greater normalized gains on the FCI, with $h = 0.35$, compared to $h = 0.18$ for the recitation classes. The results using Thornton and Sokoloff’s questions were consistent with their 1990 report, even controlling for the time spent by the two groups.
Redish et. al. concluded that MBL activities play a significant role in velocity concept formation.

The researchers further broke down their assessment by concentrating on the four questions on the FCI that deal with Newton’s third law, and calculating normalized gains for these questions. Results were tabulated for four MBL classes and six non-MBL classes (one of the classes used tutorials, but not the MBL ones). The normalized gains for the MBL classes were $h = 0.64$ and the non-MBL classes achieved $h = 0.28$.

**RQ1 in the Literature.** Redish et. al. reach a conclusion that “MBL tutorials can be effective in helping students build conceptual understanding, but do not provide a complete solution to the problem of building a robust and functional knowledge for many students.” (p. 52) At the end of the paper they state that

The Thornton-Sokoloff conjectures appear to be confirmed by a variety of anecdotes describing the success of the substitution of active-engagement MBL activities for traditional labs, and by the failure of the same equipment when used as traditional labs without the engagement/discovery component. These have not, unfortunately, been documented in the literature. It would be useful to have additional detailed experiments confirming different methods in order to build an understanding of exactly what components of MBL activities are proving effective. (p. 52)

It is to be noted that the substitution of MBL activities for traditional labs are described only anecdotaly, which argues for the importance of a quantitative study.
Thornton and Sokoloff’s Second Paper. Thornton and Sokoloff (1998) evaluated the effectiveness of an instructional program that included MBL laboratories. This paper was devoted mostly to the development of a conceptual evaluation instrument called the Force and Motion Conceptual Evaluation (FMCE). This 43-question instrument bears some resemblance to the FCI, but contains a heavier emphasis on motion graphs.

In this project, data were collected from students enrolled at the University of Oregon and at Tufts University. Only about half of the students at Oregon enroll in a laboratory, which provided the researchers with treatment and control groups. In addition to the MBL labs, students also participated in Interactive Laboratory Demonstrations (ILDs) (Sokoloff & Thornton, 1997). Assessments were delivered before instruction, after traditional instruction, after ILDs, and on the final. Students who participated in labs were shown to achieve greater gains than those who did not. But since this study did not focus on the effects of the MBL labs by themselves, laboratory effects were de-emphasized when results were reported.

Sokoloff, Thornton & Laws provide online evidence of the effectiveness of their Workshop Physics program, of which RealTime Physics and ILDs are a part (Workshop Physics project, 2001). At the referenced website, bar-graph data is presented for students who have taken the FMCE as a pretest, after lectures, and after Workshop Physics. In the kinematics graphs section, the group describes the importance of the special homework and pre-lab discussions in producing large conceptual learning gains. In the dynamics section, the importance of observing real-time impulse curves is stressed.

Other Physics Laboratory Studies. Svec (1995) looked at two groups of introductory undergraduate physics classes, one of which used MBL laboratories with the
other using traditional motion labs. The two groups were enrolled in different courses, with the treatment group enrolled in Physical Science for Elementary Teachers and the control group taking General Physics. The specific nature of the laboratory procedures used for each group was not described. Effects were measured with an instrument developed for the study, which included questions taken from several conceptual exams, including the FCI. The treatment group showed greater gains than the control group, particularly in their understanding of motion graphs.

In an unpublished dissertation titled *Comparing the Effects of Different Laboratory Approaches in Bringing About a Conceptual Change in the Understanding of Physics by University Students*, Veath (1988) conducted a study similar to the present study. She assigned three laboratory sections of students taking an introductory physics course to one of three treatment groups: traditional, intermediate, and “prediction-modified learning cycle.” Veath found significantly greater conceptual learning gains for the prediction-modified group, compared to the other two. This study was conducted before the development of the FCI, and the nature of the conceptual testing instrument is unknown.

**Cookbook Labs and Inquiry Learning**

One can find many references to cookbook labs in the science education literature. The word appears in many article titles in recent years, including *A Cure for Cookbook Laboratories* (Lochhead & Collura, 1981), *Decookbook It!* (Shiland, 1997), and *A Recipe for Uncookbooking Laboratory Investigations* (Leonard, 1991). In these and many other articles teachers are advised to “throw out the instructions” (Tinnesand & Chan, 1987), and let students devise their own method to solve problems posed by the
teacher. From a constructivist viewpoint, Pushkin (1997) gives many examples of inquiry-style questions that can be integrated into physics laboratory exercises, and asserts that lab activities provide excellent opportunities to contemplate unfamiliar concepts. He argues that “when students are regimented by lab manuals that dictate what to think, how to think, and when to think, lab activities essentially lose impact for learning.” (p. 240)

**Inquiry in the National Standards.** Authors have often cited the need to bring laboratory exercises more into line with national education standards documents as a reason for “uncookbooking” laboratories. These sentiments are quite evident in this passage from *The Liberal Art of Science* (1990), published by the American Association for the Advancement of Science, as quoted by Leonard (1991):

Thus, use of the confirmatory approach in the laboratory and in the field does not contribute to the development of strong conceptual links between the natural world and the scientific theories developed to explain and predict it. Nor does this practice leave students with an accurate view of the practice of science. Rather, it contributes to the notion that the purpose of experimentation is the verification of hypotheses rather than their refutation.

Maximum benefit can be derived from laboratory and field experiences by having students work in groups and share their ideas, perceptions, and conceptions. Group design and interpretation of laboratory work are also effective strategies for exposing the changing misconceptions. In addition, students should prepare written reports describing the rationale for the experimental design, the data, and their interpretations. (p. 87)
This passage serves well as a definition of the components of an inquiry-based laboratory, from the perspective of the broad science education community. However, one can also distinguish from “open inquiry” activities as described above, and “guided inquiry.” In guided inquiry activities, the instructor has more control over the nature of the questions and methods of investigation than in open inquiry. MBL curricula, such as RealTime Physics (Sokoloff et al., 1999), used with the treatment group in this study, generally contain guided inquiry activities. These activities contain a large number of specific directions for the student, interspersed with reflective questioning to engage the learner. Students are thereby unable to “coast through” an interactive-engagement lab because they are actively engaged in thinking, even while receiving a large amount of direction.

**Definition of Interactive-Engagement Laboratories.** The preceding discussion motivates the following definition for “interactive-engagement laboratories,” for the purposes of this study: “Laboratory procedures that actively engage the learner by the use of pertinent questions integrated into the procedure, cooperative MBL activities, and an emphasis on concept formation.”

**Cookbook Labs Defined**

Leonard (1991) provides a description of a cookbook laboratory exercise when he writes:

This student [previously described] is the victim of the overly prescriptive laboratory investigation, typical of those used in college introductory science courses. Such laboratory experiences tend to begin with the instructor explaining to the students, often in some detail, what will happen during the exercise in an
attempt to make certain that the student will carry out the exercise “correctly.” The student is then left to follow a lengthy and detailed procedure in the laboratory textbook, which will occasionally call for responses such as describing what happens with the apparatus, making a drawing, or answering a specific question in the spaces provided in the manual. The entire procedure is very prescribed, that is, the student is told what to do in a step-by-step fashion for the entire exercise. (p. 84)

In a similar vein, Grote (1998) writes “Students can usually complete so-called ‘cookbook labs’ with no understanding of what they did. Frequently, they do not form a complete picture of what happened because they focus on each step independently of the others.”

Though it is not necessary to travel back so far in history to find an example, a physics laboratory manual from the first quarter of the last century (Millikan, Gale, & Davis, 1925) illustrates a cookbook approach. In each experiment, an apparatus is described and diagrammed, and directions are given for the experiment. After these instructions, there is often a table provided for recording data. Several questions appear at the end of each experiment. This format is familiar to most of today’s scientists and educators because it is the one they used in their schooling. Examples abound throughout the last century, even into the last decade (Zitzewitz & Kramer, 1992). Given their historical prevalence, it seems likely that even today they are used in many educational physics laboratories.

**Definition of Cookbook Laboratories.** Given this background, in this study “Cookbook Laboratories” will be defined as “Laboratory procedures that follow a
‘cookbook’ approach, providing detailed instructions with no reflective questions integrated into the experimental procedure, ‘fill-in-the-blank’ data tables, and specific questions that occur after the exercise is completed.”

**MBL in Hake’s Study**

In Hake’s large 1998 study discussed previously, the 62 introductory physics courses were divided into “traditional” and “interactive-engagement” groups based on several criteria. Hake (1998b) identified seven different interactive-engagement instructional strategies, these being

1. Collaborative Peer Instruction (Mazur, 1997)
2. MBLs
3. ConceptTests (Mazur, 1997)
4. Overview Case Studies and Active Learning Problem Sets (Van Heuvelen, 1991)
5. Modeling Instruction (Wells, Hestenes & Swackhamer, 1995)
7. Other (p. 9)

Hake identified 22 separate strategies that he included in the “other” category. Each interactive-engagement course used at least two and usually more of the seven methods. Twenty-one of the 48 interactive-engagement courses used MBLs, and three of the fourteen traditional courses used MBLs. While Hake’s study clearly identifies a number of effective conceptual learning strategies, the effect of MBLs was not isolated, nor were the engagement levels of the experimental procedures identified.
Hake’s Case Studies and RQ2. In the unpublished addendum to his study, Hake (1998b) describes three case studies in which MBL labs were “grafted” onto otherwise traditional courses. Average normalized FCI gains for these three courses were only 0.26, 0.25, and 0.25. These results are close to the average gain for traditional instruction, 0.23. Hake cites several problems that may have occurred with the implementation of the MBL exercises in these low-gain courses. However, courses in the same situation at the University of Oregon and Tufts University produced very large knowledge gains, as measured by the FMCE. This discrepancy led Hake to encourage investigation of the following research question: “Can Grafting of IE Laboratories Onto Traditional Courses Markedly Increase Conceptual Understanding?” (p. 28). This question is restated as Research Question #2 (RQ2).

Student Satisfaction Surveys

Existing educational research regarding the usefulness of student satisfaction data is voluminous and complex. A broad spectrum of opinion exists as to the utility of such information, from “reliable, valid, and useful” to “unreliable, invalid, and useless.” (Aleamoni, 1981; quoted in Marsh, 1984, p. 708)

Even given this controversy, many researchers adopt an intermediate posture, and are interested in student satisfaction survey data whatever their worth may be (Greenwald & Gillmore, 1997). It is possible to embrace a position where student feedback is gathered in order to improve instruction, rather than merely using it to evaluate the teacher, its most common use (Bailey, 1983).
RQ3 was formulated in this moderate and constructive spirit. Its intent was to guide the collection of comparison data in order to provide information that supplements the main goals of the study as specified in RQ1 and RQ2.
CHAPTER III

METHODS

Population and Sample

The population for this sample was undergraduate students enrolled in an introductory physics course. This population was restricted to a sample of 52 students, who enrolled in a trigonometry-based introductory course at a small private Midwestern liberal arts college during the Spring and Fall terms of 2001. Phase I (Spring) enrollment was 25 and Phase II (Fall) enrollment was 27.

The sample included 27 males and 25 females. Their year in school and majors are summarized in Figure 3.1.

Variables and Measures

Conceptual mechanics knowledge was measured for each subject using the Force Concept Inventory. The instrument was delivered as a pretest during the first few days of class, and as a posttest on the week after all lab exercises were completed. Students were told that posttest scores would be counted toward their grades.
Student satisfaction was measured with a Likert-style survey developed for this project (see Appendix A). The survey also solicited free-response feedback.

Limitations

1. The population of all students of college-level General Physics was limited to a sample consisting of students who enrolled in two course sections that were studied in this project. This sampling procedure limits the generalizability of the results.

2. The sample size was $N = 52$, which limits the statistical power of the experiment.

3. The treatment groups were determined by enrollment in one or the other of two sets of two laboratory sections. An attempt was made to obtain equal numbers of students in each group, to enhance the robustness of the statistical analysis. The use of this convenience sample makes this study a quasi-experiment, with attendant limitations in claimed causation.

4. This study was intended to be an equal-time experiment, with the two types of experimental groups spending equal amounts of time on each laboratory exercise, one two-hour lab period for each experiment. It is impossible for absolute equality to prevail, however, given the complexity of laboratory procedures and differences among students.

5. Students in the two laboratory sections attended the same lecture periods, three hours per week. In these lectures, the instructor did not make use of any reform-based instructional activities. He used a traditional lecture approach. Studies have shown (Hake, 1998a; Hake, 1998b) that this sort of instructional strategy tends to produce relatively small gains in conceptual mechanics knowledge. So
gains experienced by the treatment groups were expected to be smaller than they would be if MBL techniques were used in conjunction with other active-learning strategies. This limitation was accepted voluntarily, in order to isolate the effects due only to the MBL laboratory procedures.

6. Conceptual learning gains were compared by measuring the construct known as the “Newtonian force concept” for individual students. This conceptual construct is large and multi-dimensional, with attendant measurement difficulties. It was measured with the Force Concept Inventory (Hestenes et al., 1992). The construct validity of this test was considered in the Review of Literature.

Procedural Steps

This study was conducted using the following sequence:

1. The course instructor was asked to participate in the study.

2. IRB approval was obtained. All IRB documents appear in Appendix E.

3. Following a suggestion by the dissertation committee, a two-member advisory panel was retained to review the content and style of each cookbook lab.

4. On the first day of class in each phase, informed consent was obtained and the FCI was administered to all students.

5. Cookbook-style laboratory exercises were written for the cookbook group and reviewed by the advisory panel. These exercises appear in Appendix B.

6. During the first nine weeks of the Spring Term of 2000, the Phase I IE group participated in interactive-engagement laboratory exercises and the Phase I cookbook group participated in cookbook laboratory exercises.
7. In the week that followed completion of the last lab, all participants in Phase I took the FCI and the satisfaction survey.

8. The experiment was repeated during the Fall Term of 2001, with neither group in Phase II participating in post-lab homework assignments.

9. After the completion of Phase II, the instructor provided written feedback about the study (See Appendix D).

Treatments

The two laboratory sections in both Phase I and Phase II experienced two different levels of engagement in their MBL laboratory procedures. One group participated in interactive-engagement exercises, which in this experiment were defined as “Laboratory procedures that actively engage the learner by the use of pertinent questions integrated into the procedure, cooperative MBL activities, and an emphasis on concept formation.” The other group participated in cookbook exercises defined as “Laboratory procedures that follow a ‘cookbook’ approach, providing detailed instructions with no reflective questions integrated into the experimental procedure, ‘fill-in-the-blank’ data tables, and specific questions that occur after the exercise is completed.”

Both treatment groups participated in the same lecture section, which met three times a week. The instructor avoided interactive-engagement activities other than MBL laboratories, in order to isolate the effect of each treatment. The same instructor taught the lecture and both lab sections.

Each phase of the experiment lasted for ten weeks. Hake compared gain data for courses at six different institutions, in which the fraction of time spent on mechanics
instruction compared to a whole term ranged from 0.6 at Harvard to two courses that spent the whole term learning mechanics (Hake, 1998a). Hake found no large differences in conceptual gains, and argued that the gain difference is robust with respect to the amount of course time spent on mechanics. The introductory physics course in this study covered mechanics for approximately 87% of the course. This duration of treatment falls within those reported by Hake.

**Weekly Schedule.** During the treatment period in each phase, the interactive-engagement group participated in the following exercises from the laboratory manual *RealTime Physics: Mechanics:*

<table>
<thead>
<tr>
<th>Week</th>
<th>Chapter</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Introduction to Motion</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Changing Motion</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Force and Motion</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Combining Forces</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Force, Mass, and Acceleration</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Gravitational Forces</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>Passive Forces and Newton’s Laws</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>Newton’s Third Law and Conservation of Momentum</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>Two-Dimensional Motion (Projectile Motion)</td>
</tr>
</tbody>
</table>

Table 3.1. Exercises in *RealTime Physics: Mechanics* completed by the interactive-engagement laboratory section.

**Selection of Chapters.** The mechanics module of *RealTime Physics* contains 12 exercises. The nine chosen above were selected for their relevance to the FCI. Omitting three activities also helped accommodate the instructor’s normal class schedule.

Activities not included in this project were Lab 8: One-Dimensional Collisions; Lab 11: Work and Energy; and Lab 12: Conservation of Energy. Since the FCI does not include questions about momentum and energy, which are the main topics covered in Labs 8, 11,
and 12, it was reasonable to expect that these three labs would not contribute to conceptual gains measured by that instrument.

Conduct of the Laboratory Activities. Each experiment in RealTime Physics has been designed to last for two hours, if no extensions are used (Sokoloff et al., 1999). Extensions were therefore omitted in order to limit time on task to two hours for each exercise. Students who did not finish exercises in two hours turned in partially completed exercises in both sections. Students in the cookbook groups were instructed to begin working on the post-lab questions when there were fifteen minutes left in the lab period.

Students worked in groups of three or two for both sections. This is consistent with the recommendations of the authors, who suggest using groups of two to four students (Sokoloff et al., 1999). Every section of each RealTime Physics lab exercise was completed exactly as written (minus the Extensions). Adequate MBL equipment allowed faithful execution of all activities.

Each experiment in RealTime Physics includes a pre-lab exercise to familiarize the students with the activity and post-lab homework that reinforces that activity’s conceptual material. The authors stress the importance of completing these activities in order to achieve high conceptual gains. They were therefore included with the laboratory exercises in this study during Phase I. The pre-lab assignments completed by the cookbook groups were written to correspond with the activities that take place in the cookbook labs and are included with the Cookbook Labs in Appendix B.
During Phase II, neither group participated in the post-lab homework, in order to provide a varied comparison set. Both groups in Phase II also completed the pre-lab exercises.

The Student Satisfaction Survey.

Given the uncertainty surrounding student satisfaction surveys, no statistical hypothesis tests were performed with satisfaction data in this study. Any such quantitative analysis will be considered to be beyond the scope of this project. Student satisfaction responses will rather be presented only as supporting information for the reader, in tabular and graphical format.

The student satisfaction survey was developed specifically for this project, using some of the suggestions of Bailey (1983), and appears in Appendix A. Upon completion of Phase II, the course instructor provided written feedback about the project, which can be found in Appendix D.

Instructor and Lab Assistants.

The instructor in this project had been teaching introductory physics for about ten years. His normal style of teaching would be described as “traditional,” and he did not try to teach any differently during the study. He was somewhat familiar with the FCI, but was mostly unfamiliar with interactive-engagement instructional methods.

During each exercise, the instructor and a lab assistant were available to help the students complete the instructional activities. During Phase I the same lab assistant was assigned to both groups, and there were two different lab assistants for Phase II. The instructor and lab assistants did not make any special effort to teach mechanical concepts during the exercises through dialogue or questioning. This posture was adopted in order
to let the written instructions in the activities speak for themselves and avoid differences that could have existed between lab assistants. This was also meant to simulate implementation of the IE activities in RealTime Physics at a school where an instructor might not be present during the lab period, or where facilitators who are unfamiliar with IE teaching methods would simply provide the instructional materials to the students. This protocol was followed closely in Phase I, but was not implemented as strictly during Phase II. During Phase II, the instructor explained that he found it very difficult to avoid instruction during the labs and did more teaching during the exercises than in Phase I.

During Phase I the instructor graded all post-lab homework and included written comments. The homework was returned to the students before the next week’s lab session. Exercises were graded and returned before the next week’s lab for all four groups.

Writing the Cookbook Labs.

The two members of the Advisory Panel both had a great deal of experience writing and conducting MBL mechanics labs on the postsecondary level. While serving on the advisory panel, one member was the laboratory manager for the department of physics and astronomy at a large university. She was also a research associate in the physics education research group at that university. The other panel member was a graduate student in the same physics education research group and a former high school physics teacher.

The Advisory Panel was provided with a set of instructions, which can be found in Appendix C. After reviewing the instructions, one panel member provided this helpful list of characteristics for the two types of labs (Plano Clark, 2001):
IE Group
1. Use of prediction questions before data collection.
2. Analysis questions throughout activities, found near descriptions of the procedures.
3. Use of MBL.
4. Emphasis on the process.
5. Emphasis on conceptual understanding.
7. Higher-order reasoning - questions asking students to compare, contrast, synthesize, and evaluate.
8. Crucial interaction between students, materials, and instructor during lab.

Cookbook Group
1. No prediction questions.
2. Summary questions after activities completed, separate from write-up.
3. Use of MBL.
4. Emphasis on results.
5. Emphasis on formal equations.
7. Lower-order reasoning - questions asking students to give facts, summarize results, and apply knowledge.
8. Interaction only between students and materials.

These lists were adopted as an amplification of the definitions, and applied to the writing of Labs #2-9. Of course, the IE labs for this project were already in existence, so the IE list did not guide their writing, but provided contrast for the cookbook list. Item #8 in the IE Group characteristics would normally apply, but the level of interaction between instructor and students during the labs was held fixed between the two groups as discussed above.

With this complete set of guidelines in place, each lab exercise went through a feedback cycle of suggestions from the Review Panel and revisions. Each cookbook lab was given the same title as its counterpart IE lab. The final versions of the cookbook labs that appear in Appendix B were then completed by the students in the Phase I and Phase II cookbook groups.
Given that no available cookbook MBL lab activities were found in published form, the cookbook labs were written “from scratch.” Some of the lab activities were modifications of procedures that had been previously used by the researcher while teaching the course. Each lab was meant to cover the same concepts as the labs in RealTime Physics, and many of the activities in the cookbook labs are similar to those in RealTime Physics, but written in “recipe” style. A few of the procedures and all of the graphics were taken from Physics Labs with Computers, Volume 1, published by PASCO scientific (1996), which is the activity guide one receives from PASCO when purchasing their MBL equipment sets.
CHAPTER IV  
RESULTS  

The Hake Factor  

Following Hake (1998a), average normalized gains (commonly referred to as the Hake Factor, $h$) on the FCI were calculated for individuals using the formula

$$h = \frac{\text{post} - \text{pre}}{1 - \text{pre}}.$$  

These individual gains are displayed on a Hake Plot in the Figure 4.1.

Figure 4.1. Hake Plot for individuals ($N = 52$) in each treatment group. Phase I included homework, Phase II did not.
Average Normalized Gains

To assess the effectiveness of the instructional treatment for each of the four experimental groups, average normalized gains were calculated for each group. The average gains are reported in Table 4.1, along with average pretest and posttest scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pre</th>
<th>Post</th>
<th>h</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cookbook I (with HW)</td>
<td>11</td>
<td>0.303</td>
<td>0.570</td>
<td>0.392</td>
<td>0.174</td>
</tr>
<tr>
<td>IE I (with HW)</td>
<td>14</td>
<td>0.298</td>
<td>0.621</td>
<td>0.471</td>
<td>0.201</td>
</tr>
<tr>
<td>Cookbook II (no HW)</td>
<td>14</td>
<td>0.343</td>
<td>0.548</td>
<td>0.334</td>
<td>0.257</td>
</tr>
<tr>
<td>IE II (no HW)</td>
<td>13</td>
<td>0.338</td>
<td>0.646</td>
<td>0.480</td>
<td>0.257</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of pretest scores, posttest scores, and mean normalized gains for the four experimental groups.

In performing the calculations above, one individual was excluded from the IE I group. This subject withdrew from the course the same week that she took the posttest FCI, and actually scored lower on the posttest than the pretest (pre = 10 correct, post = 7 correct). The instructor felt that this student was probably not taking the test very
seriously since she knew she would be immediately withdrawing from the course. This was also the only individual to show negative gain on the FCI. If this score is included in the group, the average gain for IE I becomes 0.430.

Figure 4.3 shows the Hake plot for the group averages.

Figure 4.3. Hake plot for group averages. The two dashed lines divide the allowable region into areas of high-gain, medium-gain, and low-gain. The two thin solid lines show the average normalized gains for the 14 traditional courses and the 48 IE courses in Hake’s study, which are included in the plot.

Hake (1998a) noted that the absolute value of the slope of a line connecting any point with the point (1,0) in the lower right-hand corner of the plot is equal to the normalized gain for that point. The plot can therefore be divided into three regions that
correspond to high-gain courses \((h \geq 0.7)\), medium-gain courses \((0.7 > h \geq 0.3)\), and low-gain courses \((h < 0.3)\). Two thin solid lines are also plotted that show the value of the average gains for the 14 Traditional and 48 IE groups in Hake’s data set, 0.23 and 0.48. These courses are also included in the plot. In Hake’s study, all 14 Traditional courses fell in the low-gain region. In the present study all four groups are in the medium-gain region, with the Cookbook II group near the low-gain borderline.

**T-Test Results**

Unpaired one-tail \(t\)-tests were performed on the two pairs of groups to determine if differences in mean gain scores were statistically significant. The results are presented in Table 4.2.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Mean Difference</th>
<th>Degrees of Freedom</th>
<th>(t)-Value</th>
<th>(p)-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.079</td>
<td>23</td>
<td>1.036</td>
<td>0.1556</td>
</tr>
<tr>
<td>II</td>
<td>0.146</td>
<td>25</td>
<td>1.478</td>
<td>0.0760</td>
</tr>
</tbody>
</table>

Table 4.2. Results of unpaired \(t\)-tests between the IE and cookbook groups in both phases of the experiment.

Neither of the differences were significant at the 5% level \((p > 0.05)\), though the results in Phase II are significant at the 10% level \((p < 0.10)\).

**Student Satisfaction Survey**

The student satisfaction survey appears in Appendix A. For purposes of reporting, the following abbreviations will be used.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>It was difficult to complete the lab exercises in the allotted time.</td>
</tr>
<tr>
<td>Interesting</td>
<td>My lab experiences have been very interesting.</td>
</tr>
<tr>
<td>Challenging</td>
<td>I had to work hard during the lab exercises.</td>
</tr>
<tr>
<td>Worthwhile</td>
<td>The pre-lab exercises were worthless.</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>I enjoyed the lab exercises.</td>
</tr>
<tr>
<td>Understandability</td>
<td>It was difficult to understand what the lab procedures told me to do.</td>
</tr>
<tr>
<td>Importance</td>
<td>The lab activities are the least important part of this course.</td>
</tr>
<tr>
<td>Learning</td>
<td>I learned a lot from the lab exercises.</td>
</tr>
<tr>
<td>Equipment</td>
<td>I enjoyed working with the computer-interfaced lab equipment.</td>
</tr>
<tr>
<td>Learn Physics</td>
<td>The lab procedures helped me learn physics.</td>
</tr>
<tr>
<td>Partners</td>
<td>The lab procedures made it difficult for me to work with my partners.</td>
</tr>
<tr>
<td>Thinking</td>
<td>The lab procedures made me think.</td>
</tr>
<tr>
<td>Homework</td>
<td>The post-lab homework was valuable.</td>
</tr>
</tbody>
</table>

Table 4.3. Abbreviations used for reporting satisfaction survey results.

To generate the averages in Table 4.4, responses were adjusted so that a higher score corresponds to a positive response. In other words, a high score indicates agreement with a positive question and disagreement with a negative question. Each question ranges from 1 (very negative) to 5 (very positive). Pooled standard deviations for each question are included.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Cookbook I</th>
<th>IE I</th>
<th>Cookbook II</th>
<th>IE II</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>2.000</td>
<td>2.714</td>
<td>2.615</td>
<td>2.615</td>
<td>1.255</td>
</tr>
<tr>
<td>Interesting</td>
<td>3.818</td>
<td>3.286</td>
<td>4.143</td>
<td>3.923</td>
<td>0.848</td>
</tr>
<tr>
<td>Challenging</td>
<td>3.636</td>
<td>3.071</td>
<td>3.357</td>
<td>3.923</td>
<td>1.019</td>
</tr>
<tr>
<td>Worthwhile</td>
<td>2.818</td>
<td>2.643</td>
<td>3.500</td>
<td>2.846</td>
<td>1.102</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>3.818</td>
<td>3.714</td>
<td>3.923</td>
<td>3.615</td>
<td>0.673</td>
</tr>
<tr>
<td>Understandability</td>
<td>3.091</td>
<td>3.714</td>
<td>4.071</td>
<td>3.385</td>
<td>0.934</td>
</tr>
<tr>
<td>Importance</td>
<td>3.909</td>
<td>3.769</td>
<td>3.929</td>
<td>3.615</td>
<td>0.849</td>
</tr>
<tr>
<td>Learning</td>
<td>3.364</td>
<td>3.643</td>
<td>3.643</td>
<td>3.769</td>
<td>0.911</td>
</tr>
<tr>
<td>Equipment</td>
<td>3.636</td>
<td>3.643</td>
<td>3.929</td>
<td>3.769</td>
<td>0.883</td>
</tr>
<tr>
<td>Learn Physics</td>
<td>3.636</td>
<td>4.000</td>
<td>3.786</td>
<td>3.615</td>
<td>0.807</td>
</tr>
<tr>
<td>Partners</td>
<td>3.818</td>
<td>3.929</td>
<td>4.357</td>
<td>3.923</td>
<td>0.804</td>
</tr>
<tr>
<td>Thinking</td>
<td>3.909</td>
<td>3.929</td>
<td>3.857</td>
<td>4.077</td>
<td>0.850</td>
</tr>
<tr>
<td>Homework</td>
<td>2.909</td>
<td>2.929</td>
<td>NA</td>
<td>NA</td>
<td>1.038</td>
</tr>
</tbody>
</table>

Table 4.4. Average response scores to the Student Satisfaction Survey for the four experimental groups. High scores correspond to positive responses.

Figure 4.4 presents average responses for the four groups graphically.
Figure 4.4. Bar graph of satisfaction survey responses by treatment group.

Figure 4.5 shows responses to the survey split by the type of group, IE or cookbook.
Figure 4.5. Bar graph of satisfaction survey responses by type of treatment group.

Figure 4.6 shows responses to the survey split by term, i.e. whether or not the participants also did the homework.
Figure 4.6. Bar graph of satisfaction survey responses by term, i.e. whether or not the participants also did the homework.

The opinion survey also included three free-response questions. Student responses are summarized in the following 3 tables.
<table>
<thead>
<tr>
<th>Do you have any suggestions that would improve our Phys-111 labs?</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
</tr>
<tr>
<td>Make them shorter</td>
</tr>
<tr>
<td>Include better instructions</td>
</tr>
<tr>
<td>Allow us to have the labs when we do the HW</td>
</tr>
<tr>
<td>Use activities that are less similar</td>
</tr>
<tr>
<td>Include more real-life applications</td>
</tr>
<tr>
<td>Make the labs coincide better with regular class</td>
</tr>
<tr>
<td>Make instructions less succinct/redundant</td>
</tr>
<tr>
<td>Make less reliant on computers</td>
</tr>
<tr>
<td>Make less mathematical</td>
</tr>
<tr>
<td>Don’t use this book</td>
</tr>
<tr>
<td>Make the pre-lab questions easier</td>
</tr>
<tr>
<td>Use more diagrams</td>
</tr>
</tbody>
</table>

Table 4.5. Summary of student responses to the first free-response question on the opinion survey.

<table>
<thead>
<tr>
<th>What has been your favorite part of Phys-111 labs so far?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowling ball lab</td>
</tr>
<tr>
<td>Projectile lab</td>
</tr>
<tr>
<td>Working with computer</td>
</tr>
<tr>
<td>The experiments</td>
</tr>
<tr>
<td>Fan carts</td>
</tr>
<tr>
<td>Skateboard lab</td>
</tr>
<tr>
<td>Working with people</td>
</tr>
<tr>
<td>The graphs</td>
</tr>
<tr>
<td>Hands-on activities</td>
</tr>
<tr>
<td>Using a varied approach</td>
</tr>
<tr>
<td>Tracks and carts</td>
</tr>
<tr>
<td>Easy grade</td>
</tr>
<tr>
<td>Equations</td>
</tr>
<tr>
<td>Good instructions</td>
</tr>
</tbody>
</table>

Table 4.6. Summary of student responses to the second free-response question on the opinion survey.
<table>
<thead>
<tr>
<th>What has been your least favorite part of Phys-111 labs so far?</th>
<th>Cook I</th>
<th>IE I</th>
<th>Cook II</th>
<th>IE II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too long, not enough time</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>The homework</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-labs</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Computer malfunctions</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Velocity labs</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doing the labs</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not “real-life”</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The questions</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Wordy procedures</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor integration with class</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotonous and repetitive</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Too complicated</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Motion labs</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Nothing</td>
<td></td>
<td>3</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Plugging in numbers</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>My lab partner</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.7. Summary of student responses to the third free-response question on the opinion survey.
CHAPTER V
DISCUSSION

In this chapter, each of the research questions will be considered in order. They are re-listed below for convenience.

RQ1. Are there significant differences in the conceptual mechanics knowledge gain (as measured by the FCI) for students who participate in active-learning MBL physics laboratories, compared to students who participate in equal-time exercises with cookbook procedures that also make use of MBL equipment?

RQ2. Can the use of interactive-engagement laboratories in conjunction with an otherwise traditional classroom environment produce significant gains in conceptual learning?

RQ3. How do the satisfaction and perceived effectiveness of the exercises compare for students in the two groups?

Research Question #1

Data was collected in both phases of this project in order to statistically test the following null hypothesis:

\[ H_0: \text{Average normalized gains on the FCI for the IE group were not significantly higher than gains for the cookbook group.} \]

Neither Phase I nor Phase II displayed differences in average normalized gain scores on the FCI that are statistically significant at the 5% level, with \( p \)-values of 0.1556 and 0.0760, respectively, so the study failed to reject \( H_0 \). It is to be noted, however, that the Phase II data is very close to the significance threshold. A \( p \)-value gives the probability
of the group differences being due to chance, assuming the data is normally distributed. This means that there is a 7.6% probability of this difference occurring because of random effects. The standard deviations for the two groups in Phase II are equal, 0.257, which yields a medium effect size of 0.568 s.d.

A beta value ($\beta$) is the probability of committing a Type II error, which is incorrectly failing to reject the null hypothesis. In Phase II, the beta value is equal to $\beta = 0.5912$ at the $\alpha = 0.05$ level, and $\beta = 0.4376$ for $\alpha = 0.10$. The statistical power, $P = 1 - \beta$, is the probability of properly rejecting a false null hypothesis. The corresponding power values for $\alpha = 0.05$ and $\alpha = 0.10$ are $P = 0.4088$ and $P = 0.5624$, respectively.

Lipsey (1990) has shown that $P$-values in social science experiments tend to average about 0.45.

In Hake’s study (1998a), which collected data for over 6000 students, the standard deviation for the gain scores with traditional instruction was equal to 0.14. The standard deviation for the IE group was 0.04. The standard deviation for gain scores in Phase II of this study was 0.257, which indicates quite a bit more variation than Hake’s data. This variation had a negative impact on the $p$-value in the Phase II $t$-test. If, for instance, the standard deviations for the two groups had been 0.22 instead of 0.257, the $p$-value would have been 0.048.

The $p$-value in Phase II was also hurt by the small $N$ in this study. An effect size of 0.568 s.d. would have been statistically significant ($p = 0.049$) if the $N$ for both groups had been 18, rather than 14 and 13.

The differences between groups were smaller in Phase I of the study. This may be due to the fact that the cookbook group completed the homework problems in the
RealTime Physics exercises. Perhaps the similarity between this homework and the conceptual questions on the FCI had a positive effect on gain scores for the cookbook group. Regardless of the reason for the difference, Phase II provided a more accurate comparison between active-learning and cookbook laboratory procedures, since one would not usually expect a traditional lab to include a conceptual homework component.

This study was not undertaken to investigate the effect of using conceptual homework in conjunction with laboratory exercises. It is interesting to note, however, that the added homework did not seem to have an effect on the performance of the IE groups, while it did help the cookbook groups.

It may have been more desirable to make a direct single-term comparison of an IE group that did the homework and a cookbook group that did not, since this more accurately represents typical classroom implementation of these treatments.

Given this discussion, it seems appropriate to argue that the answer to RQ1 is a qualified “yes.” The more relevant Phase II data skirts the threshold of statistical significance. This argument is made stronger by the tight controls established between the two groups. Students attended the same lectures, completed the same daily work, and spent the same amount of time in the laboratory (though it should be acknowledged that the instructor reported that he did more “teaching” during the labs in Phase II).

Are the Cookbook Labs Realistic? The validity of the results regarding RQ1 depend strongly upon the precise nature of the cookbook labs that were written for this study. Any weaknesses in these labs translate directly into a lack of generalizability of the results of this study. It is left to the reader to inspect the exercises and decide if they provide a genuine contrast to the engaging procedures found in RealTime Physics.
The cookbook labs are not the types of activities that one would expect a “real” traditional teacher to use in their educational laboratory. This is especially true since they closely follow the chapters in RealTime Physics in an attempt to provide equal content.

Several of the labs mirror the activities of the corresponding RealTime Physics chapter section by section, with the engaging portions of the activities (embedded questions, predictions, etc.) removed. By removing these engaging elements, the cookbook labs became shorter than their IE counterparts. To compensate for this, the cookbook labs included a stronger emphasis on calculations and sometimes repeated trials or used more variations in the experimental setup. Entirely different activities were undertaken in some of the labs, especially the last two.

Whether the labs followed the RealTime Physics activities or took a diverse approach to the material, the inclusion of activities depended upon the approval of the Review Panel. The nature of the labs, and whether or not they provided a true example of a cookbook lab, therefore relied on the collective opinion of the author and these two individuals.

The Instructor’s Critique. The instructor in this project provided feedback after the study was completed (Hermann, 2001). Regarding the IE labs, he said

As it was, the IE labs often felt like a set of rather unconnected experiences, which the students would hopefully put together into one conceptual framework. There was no well-defined beginning, middle and end; there were only as many experiences as the student could get through in two hours. The transitions between exercises within a lab made sense to me, but the students never noticed them and often asked about what they were doing and why.
The traditional labs also suffered from this problem, but even more so. Traditional labs usually have a beginning (deriving some equation, designing an experimental set-up, calibrating the instruments) a middle (taking data, analyzing it in graphs and/or equations, coming up with a result) and an end (calculating errors or uncertainties, writing conclusions). These traditional labs had no such parts. They were simply the IE labs without the connecting prose and questions, so there was even less to guide the student through the lab and show why they were doing what they were doing. In some cases this lack of excess verbiage actually helped the students see the “big picture” of what they were doing, but in my opinion it did not make these labs “traditional”. (p. 6)

“Traditional” vs. “Cookbook”. During the planning and implementation phases of this study, the treatment groups were labeled “IE” and “Traditional.” During the analysis phase it was decided that it would be more accurate to refer to the “Traditional” groups as “Cookbook” groups. The title of the project was also modified at this time. This decision was based largely upon the instructor’s opinion quoted above, and his comments are easier to understand if one realizes that the cookbook labs were referred to as “traditional” labs when he conducted them.

Length of Treatment. The length of the treatment in this study deviates from what one would consider “traditional.” Even reducing the twelve chapters in RealTime Physics to nine, more time may have been devoted to mechanical concepts than would be normal in a traditional course.

Given these shortcomings of verisimilitude, it should be acknowledged that RQ1 could have been answered more definitively if the quality of the cookbook labs had been
higher. In an ideal situation, the cookbook groups would have completed labs already in existence, in order to ensure comparison with a realistic alternative to RealTime Physics. Lacking this, these results should probably be considered to apply more to “cookbook” labs than to “traditional” labs, since physics educators would be likely to classify the non-IE labs in this study as the former even if not the latter.

Research Question #2

The average gain for the IE groups in Phase I and II were 0.471 and 0.480, so their pooled average gain was 0.476. The average gain for all IE groups in Hake’s study was 0.48. This similarity would seem to indicate that interactive-engagement laboratories in conjunction with an otherwise traditional classroom environment can produce gains in conceptual learning comparable to those in a classroom where the teacher uses interactive-engagement teaching methods.

This conclusion may be mitigated by the fact that the cookbook groups in this study also performed rather well. In making this comparison, it’s best to consider only the Cookbook II group, since their instructional treatment did not include the conceptual homework. Their average gain score was 0.334. This is greater than the average traditional gain in Hake’s data, 0.23. The conceptual mechanics knowledge gained by the students in the cookbook group in this particular class was above average, but the difference in the gain scores for the two Phase II groups also argues for a positive answer to RQ2.

As noted above, the duration of the treatment was probably longer than one would expect in a traditional course. The decision to graft such a long set of IE labs onto a
traditional course would depend on the preferences of implementing instructors, and it should be noted that the results of this study were achieved with a nine week treatment.

Instructors who decide to graft an IE lab onto a traditional course might also encounter problems because their lecture topics are not synchronized with the laboratory exercises, as was the case in this project. The instructor spoke to these issues by saying:

To use the IE labs effectively, I think that the instructor needs to be able to vary the way time is spent in lab and in lecture. Instead of having one two-hour lab each week and three lectures, an effective use of time would be to spend one whole week doing several of the labs, and then a few weeks without doing any labs. As it is, if the lectures continue at their normal pace, the students are well past the concepts covered in many of the labs by the time that they do the lab, and the exercises seem like a tedious chore to develop something that the students are already familiar with. On the other hand, if the lectures keep pace with the labs, then the amount of material covered must change dramatically. It may be nice to spend three weeks developing Newton’s Second Law and over five weeks with forces, but this is at the expense of a good coverage of momentum, and all mention of rotations, torque, and oscillations. To say that IE labs do a great job of teaching forces is not really fair, since I could produce students expert at almost any topic if I spent five weeks on it. (Hermann, 2001, p. 6)

Research Question #3

While the application of statistical hypothesis testing to the Likert-style survey data in this study has been avoided, one can make several generalizations about the numerical responses to the questions, as displayed in Figures 4.4-4.6.
In Figure 4.5, responses are compared by group type, IE vs. cookbook. In the differences that exist, students responded more strongly in favor of the cookbook labs in questions having to do with personal preference, including Interesting, Worthwhile, Enjoyment, Understandability, and Importance. The IE labs were rated slightly better in questions having to do with perceived effectiveness, including Learning, Learn Physics, and Thinking. These differences may or may not be generalizable, but the pattern is evident.

When comparing the Phase I and Phase II responses in Figure 4.6, even larger differences are apparent. The students who did not do the homework problems responded more positively to the labs in every question except for Importance, Learn Physics, and Thinking (which was a tie).

Qualitative observations by the instructor seem to support the patterns described above. According to him, the students who did the cookbook labs seemed to be generally more satisfied with them, especially during Phase I. This may be due to the precision of the instructions associated with these labs, which allowed the students to “cruise through” the procedures, following the recipe to its conclusion. Perhaps the cookbook labs were easier to complete, and thus more popular.

The instructor further elaborated on the popularity of the labs, explaining that the IE labs were less popular “because students in the traditional labs could at least tell what concepts they were working on, even if they couldn’t finish the lab, while students in the IE labs often had no idea what they were doing or what they were supposed to get from the lab if they couldn’t finish it.” (Hermann, 2001, p. 6)
In an informal conversation after the project was completed, the lab assistant for the Cookbook II group expressed the opinion that the lower-achieving students in her group tended to be more satisfied with the cookbook labs. She noted that high-achieving students were often frustrated by the precise mechanical instructions but that students with little interest in physics appreciated them.

**Suggestions for Future Research**

It would be useful to repeat Phase II of this study with a larger sample size. If the differences between IE and cookbook groups would be persistent, the results would have more statistical significance.

It might also be interesting to look more closely at the relative importance of the conceptual homework that is included with the *RealTime Physics* exercises. The results of this study suggest that it has little effect on conceptual learning gains for students who complete IE lab exercises, but that it does have some benefit for students who do not. Is this due to the similarity between the homework and the questions on the FCI, or does it indicate a genuine shift in beliefs about mechanical concepts? Can students in a traditional course merely complete these homework sets to produce a measurable gain in FCI scores?

The cookbook labs in this project suffered from a lack of realism. Perhaps further research could make a comparison between IE labs and labs that would truly be considered “traditional.” Are there mechanics lab exercises in existence that would provide a more realistic contrast to the IE labs? Might there be other benefits to traditional labs that are not measured by the FCI, like the ability to make quantitative
calculations? This type of study may require the use of measurement instruments other than the FCI.

Perhaps a survey could be undertaken to determine the characteristics of labs that are currently being used by physics teachers. It would be interesting to find out how common cookbook labs are, and to ascertain the actual characteristics of traditional labs that are currently being used by physics teachers.

Further research could also be conducted to measure the attitudes of the students toward the two types of labs more extensively. It would be helpful to gather qualitative data to probe these attitudes. Could it be found that preferences for one type or the other correlate with other student attributes, like educational backgrounds, attitudes toward science, GPAs, or math skills? These preferences may help explain the persistence of traditional lab procedures in educational physics laboratories.
In his famous Nike commercial series, Spike Lee’s character Mars Blackmon watched Michael Jordan in action and asked the question “Is it the shoes?” Obviously, the shoes do not make the player, but better shoes can help basketball players perform better. With the proliferation of educational physics laboratory equipment available today, a physics teacher might likewise ask “Is it the gadgets?”

Merely using MBL gadgets in an educational physics laboratory doesn’t make students learn. The ubiquitous negative references to cookbook labs in the educational literature support the common-sense notion that mindless lab procedures are less likely to produce conceptual learning. By deliberately writing cookbook laboratory procedures and comparing them to interactive-engagement procedures, this study was undertaken to acquire quantitative data in support of this seemingly obvious idea.

The cookbook labs were written to be as error-free as possible. They contained accurate information and directed the students to perform activities that worked, made sense, and illustrated mechanical concepts. They were written rigorously, but in a way that deliberately failed to engage the student while the exercise was being undertaken and lacked embedded conceptual instruction.

In a tightly controlled setting, where students covered identical concepts, used identical equipment, attended identical lectures and spent an equal amount of time in both groups, the students who completed the more engaging labs achieved higher normalized gains on a pretest-posttest FCI exam. Under these circumstances, the engagement level
of the laboratory procedures produced a difference between groups of 0.568 standard deviations, a medium effect size.

Since the students in the interactive-engagement lab groups attended an otherwise traditional physics course, it can also be argued that it is possible to have students in a traditional course participate in an interactive-engagement lab section and attain respectable gains in conceptual mechanics knowledge. These students acquired an average normalized gain of 0.476, which compares favorably with reported gains utilizing more extensive treatment regimes.

Opinion data collected in this study seems to suggest that the cookbook labs were somewhat more popular with the students than the interactive-engagement labs. These differences were fairly small, and one could argue that they are offset by larger learning gains.

By providing a quantitative argument about the effectiveness of actively engaging procedures in the introductory physics laboratory this study provides individual physics instructors with more information with which to make decisions about the nature of the laboratory exercises they provide for their students. It’s not just the gadgets, it’s how you use them that counts.
REFERENCES


Greene, R. (2000, 10/30/00). Post to PHYS-LRNR Listserv, [Email].


Hake, R. (2000, 11/10/00). Post to PHYS-LRNR Listserv, [Email].


