

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, synopsise 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.
5. Teachers assess student learning in the context of daily teaching. (Brooks & Brooks, 1993, p. ix-x)

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, $Gain_{normalized} = h = \frac{post - pre}{1 - pre}$ (where *pre* and *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).

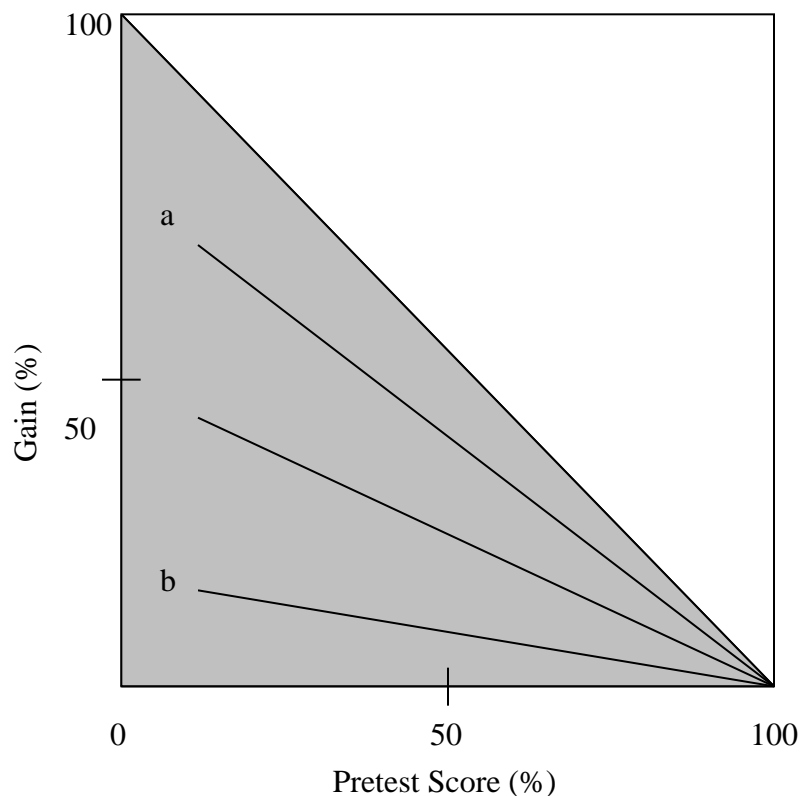


Figure 2.1. A schematic diagram of the Hake Plot. Class average gains are plotted against pretest scores. Interactive-engagement courses tend to cluster along lines closer to the top of the shaded region, with traditional courses closer to the bottom.

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 ± 0.14 (std. dev.), while the traditional courses had an average gain of 0.23 ± 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 $(2.3 \pm 1.2)\%$ 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 anecdotally, 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)
6. Socratic-Dialogue-Inducing Labs (Hake, 2002)
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.