A STUDY OF DIVERGENT TA TEACHING STYLES IN INQUIRY BASED LABORATORY EDUCATION

by

James H. Little

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF PHYSICS
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

2006

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by James Little

entitled A Study of Divergent TA Teaching Styles in Inquiry Based Laboratory Education

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

____ Date: 05/02/06

Ingrid Novodvorsky

____ Date: 05/02/06

William Wing

____ Date: 05/02/06

Dimitrios Psaltis

____ Date: 05/02/06

Michael Shupe

____ Date: 05/02/06

Keith Dienes

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 05/02/06

Dissertation Director: Ingrid Novodvorsky

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the library.

Brief quotations from the dissertation are allowable without special permission, provided that accurate acknowledgement of the source is made. Request for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgement the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ____________________
James H. Little

TABLE OF CONTENTS

TABLE OF CONTENTS –*Continued*

TABLE OF CONTENTS –*Continued*

TABLE OF CONTENTS –*Continued*

LIST OF TABLES

LIST OF TABLES *–Continued*

ABSTRACT

This dissertation is a study of the divergent behaviors evidenced by different TAs teaching inquiry based physics laboratories with minimal preparation on how to use techniques such as Socratic dialogue, wait time, and time management. The revised physics laboratory curricula, a four semester laboratory sequence, were studied over the course of two years and one of the laboratory manuals was rewritten and new techniques of TA training developed in order to align TA behavior with the ideals of inquiry based education. This revision was only partially successful, aiding TAs dramatically in improving their time management skills and use of their time, however not yielding dramatic improvements in their use of Socratic dialogue or leading questions.

CHAPTER 1

INTRODUCTION

1.1 Motivation For Study

This study was begun in order to assess the educational effectiveness of the revised physics laboratory curricula for physics and astronomy students at the University of Arizona. These revised curricula were developed by University of Arizona physics department professors in collaboration with PhysTEC Teachers in Residence. The demographics of the students studied is presented in Appendix C.

Analysis of data from the first semester of the study indicated that TA behavior had a dramatic effect on how the labs were conducted. In order to investigate how divergent different TAs' behaviors were and to assess the effectiveness of their methods of administering the labs, different observational protocols were designed to assess TA behavior.

1.2 Format of Study

1.2.1 Goals of the Lab

The labs studied comprise a four semester sequence and the pedagogical goals of the first three semesters are slightly different from the final semester, thus the labs will be described in separate sections in order to make the differences apparent.

1.2.1.1 First Three Semesters

The first semester of laboratories focuses exclusively on mechanics, the second semester on thermodynamics, vibrations and waves, fluid mechanics, and relativity, and the third semester on electricity and magnetism as well as circuit design. The courses associated with the laboratories cover the same topics as the laboratories, and in the case of the first two laboratory courses the lab is designed with the anticipation that the material examined in lab will be covered in lecture first.

According to both the authors of these labs and the author's observations, these labs are designed to, among other things, indoctrinate students in the culture of science. All of the labs have been written in such a way as to confront students with physics problems in which it is not obvious what the solution to the problem will be, nor is it clear what experimental technique should be employed to investigate the topic(s) of the lab.

Ideally, with prompting from the lab TAs and group work facilitated by whiteboard discussions, the students will be able to surmount the challenges inherent within the laboratory design. Thus the labs will force students to design and implement effective experiments in order to answer the questions suggested by the labs. The first semester of the course is designed so that the labs contain progressively less and less useful contextual information designed to help the students solve their problems as they develop a more sophisticated skill set for analyzing laboratory challenges and are able to direct more of their laboratory experience themselves.

The core idea behind this approach is that exposing students to such an environment will help students become immersed in the processes associated with doing science. If students are able to thrive in a loosely directed laboratory, then the steps leading to actual research will be easier for them to take. Since the labs empower students to arrive at their own conclusions, undergraduates taking the labs should grow in confidence and in their ability to solve problems without direction.

The majority of the labs written for the first two semesters are designed to approach material that has already been covered in the lecture, so that the students already have a mathematical skill-set with which to approach the laboratory. The belief behind this element of course design is that this will spare the lab TAs the process of helping students develop language to describe physical phenomena, and the students will instead be able to devote all of their intellectual prowess to probing the physical concepts associated with each laboratory.

1.2.1.2 Final Semester

The author of the optics and modern physics labs is somewhat more concise with the expectations that he has from laboratory work. The labs should:

1. present an opportunity for students to interact with the equipment.

2. allow students to notice and observe different experimental results illustrative of

physical principles while doing the labs. Explaining and understanding these results should comprise the real reward for the laboratory experience.

3. not be comprised of a set of objectives, rather should present an opportunity for the students to explore either:
    (a) the intuition that students have developed in life.
    (b) the intuition that students have developed from doing homework.

The fourth semester laboratories contain suggestions for interesting and useful experiments for students to investigate and do not obscure these experiments by embedding them within contrived premises as the first and second semester labs do. All four laboratories ask students to design their own experiments, however the direct suggestion for experiments contained in the fourth semester lab aid students in getting started faster and with less frustration.

In a conversation with the author of this dissertation the author of the fourth semester laboratories stated that he was dubious about investing all responsibility for organizing the lab in the TAs, preferring to embed instructions within the lab manual to guard against TA errors stymiing students. The author of the first through third semester laboratories is a theorist by training, and the author of the fourth semester labs an experimentalist, and the different backgrounds of the authors is illustrative of the differences between the two labs.

1.3 Criteria for Analysis and Revision of Laboratories

All of the observation techniques developed were used to assess the effectiveness of the laboratories. A basic assumption was made that the laboratories are more effective when TAs interact with a large number of the students, TAs use Socratic dialogue to engage students, all the students in the room are engaged in doing the experiment, and students are formulating and sharing useful conclusions about the experiment in lab.

Observations were completed in the spring and fall semesters of 2004. In spring 2005 the second semester laboratories were rewritten by the author of this dissertation to conform to the following criteria:

1. increase the number of discussion questions and relevance of discussion questions used in the lab.
2. increase the number of student presentations during discussion sessions and discourage TAs from working out problems for students.
3. standardize the teaching schedule for teaching uncertainty analysis and ensure that the basics of this analysis are taught in all classrooms.
4. expose students to the culture of science through research in physics history.
5. eliminate TA misconceptions about the laboratory through careful explanation of theory behind lab.
6. control TA lecture length and content.
7. help TAs administer classrooms through Socratic dialogue and leading questions.
8. standardize grading between classrooms.
9. reorganize the way time is used in classrooms so as to promote closing presentations on student work.

1.3.1 Observation Protocols

Various different observation techniques were used to study TA behavior throughout the course of this study. Briefly, these protocols include:

- SATIC and ROP, which were used to analyze the content and length of TA lectures, as well as the activities TAs engaged in while lecturing.
- IR:SQ which was used to track how TAs interacted with students during small group discussions.
- Time-Motion studies which were used to assess how long TAs interacted with groups, how much of the TA's time was spent interacting with groups, and what percentage of time TAs initiated contact vs. responded to requests for help.

1.3.2 Student Interviews

Two different types of interviews were conducted in this study, group interviews and individual interviews. Group interviews were used as a method of assessing the feeling of a large number of students about the labs. Individual interviews were used to assess the effects of the labs on student understanding and, to some extent, their attitudes.

1.3.3 Written Formative Assessment

The revised labs often addressed a single topic or small number of topics, and these topics ranged from learning a particular mathematical technique to a particular concept in physics (such as conservation of energy). The students' understanding of these topics was assessed using written quizzes administered in the first ten minutes of the following lab session.

1.4 Format of Remainder of Dissertation

The remainder of this dissertation is organized in the following matter: chapter two is a literature review which contains descriptions of the relevant articles that were used to inform the study of the TAs and the revisions of the laboratory manual. Chapter three is an overview of the observational techniques, written tests, and interviews used to assess the effectiveness of the laboratories. Chapter four contains the majority of results, with the exception of detailed quiz data, which is included in the document "Study of and Recommendations for Inquiry Based Labs in the University of Arizona Physics Department." Finally, chapter five contains conclusions generated from the study, reflection about the validity of the results, and general recommendations for the overall labs.

The document "Study of and Recommendations for Inquiry Based Labs in the University of Arizona Physics Department." contains analysis of each individual lab studied over the course of this project, quiz data attendant to laboratory, and itemized recommendations for methods of revising laboratories. It also includes the bulk of the rewritten second semester laboratory manual, so as to provide an alternate viewpoint on how these laboratories can be implemented.

CHAPTER 2

LITERATURE REVIEW

This section is intended to provide a thorough background on issues in physics education, acquaint the reader with the research ideas which ground this dissertation, and to help provide an understanding of education research and the context in which it occurs.

This chapter opens with a description of two of the most important cognitive theorists in the world, Vygotsky and Piaget, and then moves into a discussion of the field of education that has grown up around their work, constructivism. Constructivism is significant because a strong justification of inquiry based education arises from the argument that students are better able to develop robust understanding if they are constructing their own mental picture of how a phenomena occurs. The students can do this most effectively if they develop understanding through their own explorations, which is inquiry based education.

A brief history of Physics Education Research (PER) is included next, and then a synthesis of various research results relevant to laboratory effectiveness is presented in order to set the context for laboratory assessment and reform. Finally miscellaneous topics such as the definition of wait time, active engagement techniques, criteria for conducting research in social sciences, and descriptions of interview techniques will be presented. These topics are important for understanding the context and formation of this dissertation, however do not fit neatly into any of the previous categories.

2.1 Cognitive Theorists

The study of the cognitive processes associated with education is a thriving field with many notable figures. The researches most closely associated with PER have largely drawn their epistemology from the writings of two notable figures in the field: Piaget and Vygotsky. It is necessary to discuss the theories of these two luminaries in this dissertation so that readers can develop an understanding of the tacit assumptions that social scientists working in PER often make in analyzing how students learn.

Piaget and Vygotsky were among the first educational theorists to use carefully-controlled studies to test their theories. While many cognitive theorists have developed new theories since Piaget's and Vygotsky's deaths the work of these later theorists still often falls into the category of either social constructivism (associated with Vygotsky) or personal constructivism (associated with Piaget), and thus draws deeply on the work of either Piaget, Vygotsky, or both.

2.1.1 Theories of Jean Piaget

There is a story told of Piaget that helps to illustrate his techniques of experimental design and the methods by which he categorizes learners. Piaget waited until his first child was largely accustomed to receiving nourishment from bottle-feeding, and then he would take the bottle, rotate it 180 degrees about a non-symmetry axis, and present the bottle to the child and watch him hit his face with the wrong end and cry, presumably because he had just banged the flat end against his face and had not made contact with the expected nipple.

Piaget continued conducting this experiment until the child learned that the bottle had been rotated and then rotated the bottle back on his own. This indicated that the child had developed a new level of spatial awareness, which Piaget would note then try to categorize both at what age this development came about and what the precursors to the rotational ability might be. Later in life he would conduct similar experiments with a large number of children so as to make his conclusions generalizable to a larger population.

The working theory that Piaget developed was that "l'intelligence organize le monde en s'organisant elle-même" [1, p. 400] or "intelligence organizes the world by organizing itself." Piaget theorized that human beings interact with the world through mental constructs called structures, which develop ever-increasing complexity as people become older and more experienced. Today many of his theories have been expanded by other social scientists, and thus the synopsis of his work presented here has been partially informed by more recent researchers [2] as well as Piaget's original findings [3].

According to Piaget's theories there are four key factors that aid in the development of new structures: experience, maturation, social transmission, and equilibrium.

1. **Experience** comes in one of two forms. Children either directly experience the properties of objects (such as size, color, taste) or else they have experiences with

sorting objects by organizing them according to their extrinsic qualities.

2. **Maturation** occurs as the nervous system grows, and it is independent of a person's environment. Piaget believed firmly that children were not equipped to develop certain structures until they reached a certain age.

3. **Social Transmission** of information is required to allow children to place their developing structures in a context that allows them to effectively relate to their society.

4. **Equilibration** is the mechanism through which children assimilate information learned from their own experience, their developing brain, and society at large. Equilibration is thus the true mechanism through which structure formation occurs, and Piaget asserted that it is an entirely individual achievement, accomplished through reflection and individual effort.

Thus, Piaget imagined that the first three factors listed above provide all of the stimulus and contextual information necessary to aid children in structure formation, and that equilibration is the process by which children form their structures through internalized trial and error. By stepping through this process again and again, children go through many stages of structure formation. Depending upon which structures were being built at the time, Piaget categorized children as belonging to one of four different stages, which most children entered about the same time, provided they are maturing in a healthy developmental environment:

1. **Sensory-Motor** (Birth - 18 to 24 months)

2. **Pre-Operational** (18 to 24 months - 6 to 8 years)

3. **Concrete Operational** (6 to 8 years - 11 to 12 years)

4. **Formal Operational** (11 to 12 years - death)

2.1.1.1 Sensory-Motor

The structures that are built while a child is in the Sensory-Motor period are called "schemes." These schemes are not well-developed structures, and they amount to children learning coordinated actions. Some example schemes are the ”thumb-sucking scheme,” the ”grasp-object scheme,” and the ”scream-for-attention scheme.” Structure formation at this level is highly limited and not generally relevant to learning in a college laboratory environment, and thus will not be further explored.

2.1.1.2 Pre-Operational

When children enter the pre-operational developmental stage they develop structures called “semiotic functions”, in which objects can represent other objects. This development of schema to allow symbolic thought often occurs during play, in which objects can represent things which they are not. An example of the use of semiotic functions during play would be a child having a box represent a spaceship or a stick represent a gun. The pre-operational stage is not relevant to this study as the prerequisites to take a course on physics are such that all subjects must have passed out of the pre-operational stage in order to gain entrance into the class.

2.1.1.3 Concrete Operational

The concrete operational stage occurs when children begin to construct “operational structures,” which are structures that have been built out of what appear to be smaller structures. An operation is thus an action upon an object which is constructed out of a time-ordered use of different structures largely learned in the pre-operational stage.

Piaget stipulated that all operations need to be reversible, hence examples of operations would be adding, subtracting, and rotating. Activities such as writing, talking, or sleeping are not operations, because they cannot be reversed. An operation must be able to be done, and then undone, so that children can fully assimilate the usefulness of an operation and organize it properly within their repertoire of tools.

Concrete operational structures include classifications, ordering and seriation of objects; formation of topological, projective, and Euclidian spatial structures; additive and multiplicative structures; and one, two, and three dimensional measurement. It is in this time that most of the tools that human beings use to manipulate their environment are developed due to the ability to string together several different operations in a single structure.

An example of an operation would be the steps that a student can learn to internalize differentiation of a power-law equation such as $\frac{d}{dx}x^n$. As a student is learning differentiation, possibly cramming for a test, he/she is likely to construct an operation to handle differentiation that reflects little understanding of the processes of differentiation, but is still adequate to answer many problems. An example structure for differentiation would be:

1. Put the number in the exponent in front of the answer.
2. Write down the variable that is differentiated.
3. Write down the exponent minus one as the new exponent of the variable.

This operation is perfectly adequate for calculating $\frac{d}{dx}x^n = nx^{n-1}$, and can be constructed while the student is in the concrete-operational thought stage, but shows little understanding of differentiation. Almost all human beings have a large set of operational structures that they use to get their day-to-day business accomplished, though the type and number of structures developed varies widely from profession to profession.

2.1.1.4 Formal Operational

Formal operational thought can be achieved when children learn to think abstractly about objects, and to make inferences from the information available to them. In theory, during the formal operational phase of thought, students will learn how to use proportionality to analyze objects, will begin to understand and be able to draw inferences from correlations, will be able to construct combinatorial systems, and will be able to analyze systems from multiple frames of reference.

Most problems in physics require the use of formal operational thought. For example while one could memorize the steps for doing a Lagrangian mechanics problem and execute those steps without having reached formal operational thought, the ability to interpret the results of such a problem would certainly involve formal operational thought. Indeed, all transformation from mathematical language to physical description is grounded in this stage of cognitive development. Other characteristics of formal operational thought include the ability to separate ideal characteristics from actual characteristics of a situation, person, or thing. Formal operational thinkers are able to reason logically, and to make clear plans for solving problems.

2.1.1.5 Criticism of Piaget

While Piaget's work is regarded as being vitally important to the modern field of education, he has had several vocal critics [4] in recent years. Piaget used clinical studies of individual children to establish the framework of his theories, but he never used experimental and control groups to see if the very process of interviewing and studying the children had an effect on them. Piaget never studied individual differences in intelligence, binning all children of a certain age category into the same grouping. Piaget never studied children in social environments, to see if their behavior differed when surrounded by a peer group. The interview protocols that he used were not designed to give children help in terms of the level of language and the choice of situations used in the interviews. Piaget also never separated the results of his social experiments by gender, tacitly assuming that males and females both pass through the same developmental stages at the same times.

The most serious criticisms of Piaget's models target the assumptions that children occupy certain stages at certain ages. While the preoperational and sensory motor stage are not directly relevant to a document analyzing college-level education, the controversy surrounding the years in which people occupy formal operational and concrete operational stage certainly are. In recent years, a number of studies have questioned whether students move to what Piaget would describe as formal operational thought at age 12, or if a large body of students either never transition to formal-operational thought or else do so much later in life.

2.1.1.6 Piaget in a College Context

An experiment conducted by McKinnon and Renner [5] studied the effect of putting a population of college freshman through a course designed to transition them from concrete operational thought to formal operational thought. Their experiment first established that a large number of students were apparently operating below formal operational thought before any interference was attempted. Two groups of students were then studied, one of which (the experimental group) was exposed to a class designed to promote transition from the concrete operational stage to the formal operational stage. The other group was allowed to continue with their normal coursework without any special training, and then both groups were retested at the end of the semester. McKinnon's results are summarized in Table 2.1.

		Pre-Test Numbers			Post-Test Numbers		
Group	Stage	Females	Males	**Both**	Females	Males	**Both**
Experimental	Formal	4	11	**15**	14	16	**30**
	Concrete	38	16	**54**	28	11	**39**
Control	Formal	4	14	**18**	7	17	**24**
	Concrete	32	12	**44**	29	9	**38**

Table 2.1: A Comparison of the Growth in Logical Thought Processes of the Experiment and Control Group by Number of Students

McKinnon's results indicated that some 76% of the students entering college were operating at the concrete operational level of ability instead of the formal operational level. This indicates that a large number of the college students evidence inability, probably arising from their developmental environment, to think abstractly about problems. The study that McKinnon conducted was also interesting in that he showed that college students can be accelerated into formal operational thought by training in a course. Approximately 28% of the concrete-operational students involved in the experimental class transitioned to formal operational by the end of the year. In the groups with no intervention, only 14% of the students transitioned, indicating that the course was of significant benefit.

While little to no data exists on the cognitive level of physics students, McKinnon's study implies that caution should be taken before it is assumed that physics students are capable of abstract thought. A population of students that is taking a physics class and has yet to reach the formal operational stage would likely fail the course due to an inability to evidence the abstract thought that allows, say, a word problem to be transformed into mathematical form.

2.1.2 Theories of Lev Vygotsky

Vygotsky's theories as they are currently accepted is a superposition of his own work and the work of those who followed him. His relatively short lifespan (1896-1934) and the fact that he only spent the last ten years of his life developing his educational theories means that his ideas were not as well fleshed-out as those of Piaget at the time of his death. This section is thus a superposition of Vygotsky's own words [6] and how his work is viewed years after his death [7].

Vygotsky's work began in the social context of the U.S.S.R., which was an intellectual environment in which he appeared to thrive. It is clear from his writings that he had access to at least the work of Piaget, and possibly other western authors, but the original inspiration for his theory of cognition came from the German psychologists, Wolfgang Köehler and Kurt Koffka, who invented Gestalt psychology. Vygotsky's theories reflect the influence of psychology as his work is full of careful reasoning about potential cause and effect in human development, making it distinct from Piaget's rigid categorizing of human abilities.

Vygotsky, like Piaget, believed that children develop structures while learning, however Vygotsky saw structure creation as producing more complicated structures with maturation. These later developmental structures transform the lower structures, thus maturation is not a process of restructuring at the basic level, but of transformation from above. This conflicts with Piaget's view that structures remain basically static and are organized in different ways as children mature and more structures are added.

This interpretation of structure formation and development is significant, as Vygotsky developed a theory in which social interaction can aid in structure development. This placed him in marked contrast with Piaget, who felt that structures must be developed entirely by individuals. Vygotsky agreed with Piaget that certain problems were beyond the reach of children at certain ages, however he noted that children could be guided to solve these problems with appropriate levels of help. When a problem is beyond a child's normal facilities, but that child can solve the problem by working in a group or though prompting, then the problem is considered within the child's Zone of Proximal Development (ZPD). Vygotsky's essential claim is that intellectual maturation occurs at a more rapid rate as long as students are operating slightly above the level at which they can comfortably perform. In this fashion students are consistently exposed to an environment in which they are forced to perform above their level (which Piaget would contend is impossible).

Vygotsky's work stressed the importance of play for young children and group work for older children, as this would aid in raising an individual child's ZPD and in structure formation. Since both play and group work expose children to social interactions that help them understand the appropriate methods of interacting with their environment, they have the double benefit in that through cooperation, a problem can be solved that could not have been solved individually. In principle, all the participants in a group will have the opportunity to observe the higher level structures evidenced by their peers. This will enable the slower members of the group to better develop their structures through observation and imitation, and will aid the higher-functioning members of the group by socializing them and allowing them practice in self-examination by demonstration of structures to their peers.

Vygotsky's major publication was *Thought and Language* [8], and the majority of his ideas deal with the subtle interplay between cognitive development and the role of language. He believed that the social interactions experienced as children developed would eventually become part of the child's psychology in a process called internalization. This process of internalization is only possible because of language, which makes it possible to express and receive complicated thoughts and both reflects and shapes the society to which an individual belongs.

Vygotsky's theories were not translated into English until 1960 and not widely accepted by western educational theorists until approximately 1980. While the essence of his theories are grounded in Gestalt psychology, the social orientation and emphasis on cooperative learning was looked on with great suspicion by educators in democratic countries. In Russia itself Vygotsky was scrutinized closely because his theories on human development were highly opposed to Pavlovian psychology, which was in vogue with the Russian Communist Party at the time. He is today regarded one of the most important cognitive theorists in the United States, and the dominance of group work in high school, middle school, and even college is due in no small part to acceptance of his theories.

2.1.2.1 Vygotsky in a College Context

The theories of Vygotsky are generally applied to young children, as this is the context in which he performed most of his research. Vygotsky is extremely important to college teaching in the sense that his work is the foundation upon which social constructivism (to be discussed later) rests, however the terminology and exact specifics of his cognitive theory are generally not applied to college level courses.

However, one of Vygotsky's ideas has recently worked its way into literature focusing on college students due to the work of Lerch [9], and that is fossilization. Vygotsky first wrote about fossilization in *Mind in Society*, [6] describing it as a mechanism through which learning is lost but the attendant behaviors associated with prior learning remain.

This was noted in the context of college classrooms in which students displayed weak knowledge of algebra despite previous exposure to it. Students studied by Lerch displayed confidence that they possessed the skills to work elementary algebra problems, such as distribution of terms, but were unable to complete these relatively simple problems when put to the test.

The proposed treatments involved either presenting the college students with tasks that do not immediately appear related to the concept being reviewed and reviewing concepts through tasks that are completely new to students. These two techniques are employed to keep students from blithely assuming that they are knowledgeable on a subject and dismissive to the activities when, in fact, they have lost the skills necessary to complete various tasks.

2.2 Constructivism

Education research, both in general and as applied to the sciences, has been deeply influenced by the rise of a theory called constructivism, which deals with analysis of the methods by which students construct their knowledge. Constructivism is born out of the work of cognitive theorists, especially those of Piaget and Vygotsky. Piaget's theories have formed the basis for a subset of constructivism called personal constructivism, and Vygotsky's theories have helped to form the core of social constructivism. The most effective and visible advocate for constructivism was surely Rosalind Driver who, until her death in 1997, published numerous articles examining the underpinnings and effectiveness of constructivism, including the article [10] this section is largely derived from.

A more sophisticated definition of constructivism holds that knowledge is never directly transmitted from teacher to learner, rather teachers provide visual, verbal, and other sorts of cues upon which students construct their own notions of how the phenomena or theory under discussion works. Different methods of teaching can assist students in constructing knowledge more directly and some methods of instruction, such as lecture, have little ability to help students construct their knowledge.

Because constructivism is a very broad field it encompasses several different intellectual approaches. A few of these, taken from Driver's 1994 article [10], are as follows:

1. Carey [11], Carmichael [12], Pfundt, and Duit [13] present constructivism as being the process by which students construct informal theories about natural phenomena, and assign informal definitions to terms.
2. Piaget's [14] own work on constructivism focused on learning as resulting from a learner's personal interactions with events.
3. Edwards & Mercer [15], and Lemke [16] analyze knowledge construction as resulting from students becoming conversant with the culture of science through discussion with others.
4. Rogoff and Lave [17] view constructivism as occurring as a result of apprenticeship in scientific practices.
5. Johnston & Driver [18], Scott [19], Asoks, and Embertson [20] present ways in which students' informal knowledge is utilized when thinking about scientific concepts and methods used to teach science in classrooms.

No one educational viewpoint is appropriate to apply to all situations at all times, and very likely all of these different perspectives have greater or lesser applicability depending on the classroom situations. For example, Rogoff and Lave's theories could be useful in describing the relationship between a Ph.D. student and his/her advisor but their theories would have little applicability in a large lecture classroom. These different groups of constructivists are presented and summarized for the purpose of highlighting the divisions present within the constructivist movement as a whole. Few education researchers living today would not identify themselves as being constructivists, however the movement is large enough to encompass different, and sometimes contradictory, ethos. There exist two broad categories of constructivists, which will be reviewed here to aid in understanding the differences and tensions between the two sides of the field.

2.2.1 Personal Constructivism

Piaget is the undisputed founder of the philosophy of constructivism, hence personal constructivism is the first iteration of a version of constructivism. The basis from which Piaget constructed his approach has already been discussed, however personal constructivism has grown beyond his initial ideas. Here personal constructivism is presented both as Piaget envisioned it, and as Driver and her collaborators reinterpreted it in the 1980s and 1990s [10].

Piaget's work mainly focused on how individuals interpret their environment by developing logical structures and operations, thus the pedagogy grounded in his theories focuses on assisting structure formation. Piaget himself was relatively silent on the subject of structure formation, believing that it was primarily an interplay between developmental and environmental factors. An environment in which experience is abundant, maturation occurs, and social transmission provides context will allow a student to equilibrate these three stimuli and engage in structure formation. Children's intellectual development is linked intrinsically to their age, and their environment could be unsuitable for structure formation in the sense that the environmental input could be insufficiently rich to allow for structure formation and modification. Piaget thus laid out a theory in which development could clearly be stunted, but he never focused his research on finding methods to accelerate structure formation or development.

Driver worked on something that could be called personal constructivism, but would perhaps be better referred to as domain specific constructivism. Her approach has a great deal in common with the theories of Piaget, however it focuses on the acquisition of domain specific knowledge (mainly the domain of science) as opposed to structure formation and development. In short, her theory is that individuals develop knowledge schemes that represent their current understanding of a field. Learning occurs when the schemes change through a process of disequilibration (as opposed to the process of equilibration through which schemes are established).

This disequilibration requires internal conflict that results from presenting students with scenarios and examples that cause them to question their own current schemes. This puts the teacher in the role of both providing experiences to students and then encouraging their reflection. A good teacher has the capability to induce cognitive dissonance in students and to present students with abundant information that allows them to redesign their schemes in ways that conform to modern scientific theories.

Driver's approach to science teaching involves establishing students' prior knowledge and carefully constructing tasks that result in the students engaging in self-examination of that prior knowledge. In order for constructivism to be effective these tasks must be designed in such a way that students do not develop a new set of schema that are not compatible with modern-day science, so the teacher must often assess student learning and schema construction.

2.2.2 Social Constructivism

Social constructivism owes much of its organization and language to the work of Lev Vygotsky and it is his work, and his successors, that will be presented here [21]. The difference between social constructivism and personal constructivism is largely one of emphasis. Personal constructivism stresses the benefit of personal physical interactions with the world that allow students to construct their own understandings. Social constructivism stresses the importance of the introduction and understanding of symbols, with the idea that symbols are entirely socially constructed objects.

Vygotsky, unlike Piaget, did develop a particular pedagogical viewpoint about the appropriate environment for children. Vygotsky considered play and social interaction to be absolutely vital to early childhood development and that when children entered a formal learning environment they should be pushed to develop. His pedagogy involved pushing students out of their ZPD using a technique called scaffolding.

In scaffolding, a teacher sets a student or group of students to a task which is slightly beyond their current abilities. The level of the learners' participation is moderated by the teacher, who may render assistance in solving the problem if it is deemed too difficult. The technique is referred to as scaffolding because teachers are supposed to gradually but constantly reduce the amount of help they give students. In this way students are first shown how to do things, then helped to do things, and finally (theoretically) are capable of completing the tasks on their own.

2.3 Beginning of Research on PER

2.3.1 Context of Field

This section deals with the early history and development of the field of physics education research (PER). The field of PER is not large in comparison to many branches of physics or education, and its history is still relatively short, but this is by no means a comprehensive overview. The reader is invited to view this segment of this dissertation as a subjective reporting on PER that is meant to help those with no knowledge of the field understand the basic premises that have developed as the field has evolved.

PER is a field that has grown due largely to professors in physics departments undertaking educational research. The language of statistics that is spoken in education articles use methods such as “t-tests,” “p-values,” and their ilk not part of the terminology of most physics professors. Since this has always been the audience that physics education researchers are trying to reach avoidance of educational terminology helps PER professors to communicate to the physics community, but makes communication with education departments difficult.

Through study of the effectiveness of physics instruction people working in PER hope to describe effective new curricula and to convince physics teachers to use their innovations. Communication with the educational community is thus a distant second priority, which while understandable, has also caused PER to evidence a narrowness of view where the same few ideas and instruments are used again and again without intellectual influx from the larger community of educational researchers.

Articles dealing with the study of the educational effectiveness of physics courses have been published since the early part of the 20th century, but it is difficult to say when the field of PER came into existence. Lillian McDermott of the University of Washington began publishing papers describing systematic investigations of students’ misconceptions as early as 1980, however PER is most visible to physics faculty in the form of various diagnostic exams that have been developed to assess students’ conceptual knowledge.

The diagnostic exams traditionally used in physics are generally multiple choice tests that focus on a particular subfield of physics. The questions chosen are usually straightforward, conceptual in nature, and can be solved without recourse to mathematics. For these reasons diagnostics tests are frequently used to assess if classroom instruction has had any positive impact on students understanding of physics.

2.3.2 Diagnostic Exams

In 1984 Ibrahim Abou Halloun and David Hestenes published a test called the Physics Diagnostic Test (or PDT) [22]. The PDT is a thirty-six item multiple choice test that contains mechanics questions that students with a solid understanding of Newton's laws should be able to answer relatively easily. The simple and straightforward design of the test leads to the conclusion that any student who received a good grade in an introductory physics course should be able to respond correctly to the items.

Halloun and Hestenes used the test as a pretest and posttest over the course of three years in order to probe student understanding of physics. Early versions of the test were not multiple choice, and the students were allowed to write their answers freehand. The most common incorrect responses were used as answer choices in the final version of the PDT, based on the assumption that the most common incorrect answers would correspond to innate student misconceptions.

The authors used the PDT to show that knowledge gained by students under conventional (non-inquiry based) physics instruction was entirely independent of which professor had charge of the class. They further concluded that conventional instruction was ineffective in significantly improving performance on the PDT. Analysis of their results revealed a correlation between good performance in the course and the PDT, however even the "A" students were only able to score an average of 75% on the Physics Diagnostic Test as a posttest grade, and their pretest was, on average, 63%, indicating only a 12 % grade increase. Other grade levels showed similar pretest-posttest gains, but much lower starting grades, as shown in Table 2.2.

Grade	Number	P.D.T. pre-test	P.D.T. post-test
A	31	63%	75%
B	61	55%	67%
C	66	47%	62%
D	25	43%	56%
E	9	38%	46%

Table 2.2: Physics Diagnostic Test Results Pre and Post Instruction

The initial data set for the Physics Diagnostic Test was fairly small, however it stimulated thought, conversation, and served as a body of evidence that physics education researchers could use to indicate that conventional lecture based instruction was ineffective.

Hestenes was not satisfied with the PDT and, in 1992, published two new instruments meant to entirely replace it. The first, the Force Concept Inventory (FCI) [23] was designed to test students' conceptual knowledge of physics. The second, the Mechanics Baseline Test (MBT), [24] was designed to test students' ability to apply mathematical knowledge to physics problems.

2.3.3 Impact of Diagnostic Exam

The main paper reporting the results of the use of diagnostic instruments was a comprehensive survey of over six thousand physics students (52 introductory physics classrooms, 14 lecture based and 36 using interactive engagement techniques such as peer instruction) spread across the country [25]. The data considered were a mixture of results from the PDT, the MBT, and the FCI, depending on the level of course and choice of particular instructors. Hake choose to couch his results in terms of the average normalized gain $\langle g \rangle$ between pretest and posttest, defined as:

$$\langle g \rangle = \frac{\%(post) - \%(pre)}{100\% - \%(pre)}$$

Hake defined "high-gain" courses to be those for which $\langle g \rangle \geq 0.7$, "medium gain" courses for which $0.7 > \langle g \rangle \geq 0.3$, and "low gain" courses with $\langle g \rangle < 0.3$.

Hake's research revealed that all 14 of the traditional courses fell in the low gain region, seven of the 48 interactive engagement courses fell in the low gain region, and the remaining 41 courses fell in the medium gain region. No course participating in the study was able to achieve the high gain region. In his microanalysis of results Hake noted that of the 36 courses that achieved $\langle g \rangle > 0.6$, 48% were taught by professors who had published in physics education research, therefore his experiment may be biased towards sampling professors who are heavily involved in PER. He is quite clear that he included all data available from professors who taught traditional courses and were willing to allow their classes to be analyzed.

The validity of the FCI was analyzed by Henderson [26] who concluded that students appear to take the FCI seriously, even though it is normally given as an ungraded test. He also did a statistical study comparing the effects of giving it or not giving it as a pretest, finding that two years worth of students (class size not reported) posttest scores did not seem to be significantly influenced by taking the FCI as a pretest. This result helps to lend Hake's $\langle g \rangle$ factor greater validity.

Recently, Hake's study was reexamined by researchers who administered, along with the FCI, a test called Lawson's Classroom Test of Scientific Reasoning as a pre and post test [27]. This exam is a reformatting of Piagetian tasks to test whether students are in the formal operational or concrete operational stage. The authors made a strong argument that increases in normalized gain are quite possibly due to students transitioning from concrete operational to formal operational thought. This indicates that elementary exercises in proportional reasoning and other teaching techniques [28] designed to aid transference may be far more valuable for students than the use of active engagement techniques in classrooms.

The Force Concept Instrument is far from the only multiple choice diagnostic tool in use in physics today. These instruments have been developed for Electricity and Magnetism [29], Circuit Design, Energy (including heat), Quantum Mechanics, and so on, however none of these exams have been distributed as widely or studied as carefully as the Force Concept Inventory has.

2.3.4 Negative Reaction to the Force Concept Inventory

Perhaps the most telling criticism of these conceptual exams is that while low performance is clearly an indication of a lack of conceptual understanding of a subfield of physics, it does not necessarily follow that high performance indicates strong conceptual understanding on the part of the test-taker. The tests are multiple choice, a format notoriously prone to being vulnerable to good guessing strategies. Furthermore the instructor of a class using interactive engagement techniques will be fairly familiar with the questions on the exam and may use questions very similar to those on the exam as examples or on homework to assess their students' conceptual knowledge, and thus prepare students to perform well on the FCI because of familiarity with the questions rather than deep conceptual understanding.

This question was partially addressed by Steinberg and Sabella [30], who analyzed student performance on FCI questions and identical questions placed on tests scattered throughout the semester. Their results suggested that, for certain questions (those related to Newton's third law), students seemed capable of either correctly answering the FCI questions and incapable of answering the same questions on a test or vice versa. They noted that the FCI has no predictive power about students' views and attitudes about science, reasoning abilities, or problem solving skills.

The FCI was analyzed by Heller and Huffman [31] who noted that Hestenes and Halloun claimed that the FCI was representative of six different subfields of mechanics (kinematics, Newton's first, second, and third law, superposition, and different kinds of force), but their factor analysis of the FCI failed to resolve six distinct factors. The factor analysis indicated that college and high school students' answers are suggestive of at least 9 or 10 different factors present in the test. Following this article Hestenes and Halloun published a reply [32] and Heller and Huffman a counter-reply [33] that resulted in little dilution of the point that many of the claims that have been put forward for the FCI have been done without rigorous supporting statistical analysis. This further calls into question the value of using the FCI as a tool for classroom analysis and calls into question the wisdom of relying as heavily upon it as the PER community has to date.

2.3.5 Cataloguing of Misconceptions

The work of the University of Washington group began at roughly the same time as Hestenes and his collaborators were first beginning to publish their multiple choice tests. McDermott, the head of the Washington collaboration, is a constructivist who predicated her research on the idea that in order to properly understand how to teach students it is necessary to first assess how current methods of teaching are effecting their knowledge states and what their initial knowledge states are.

The research generated at the University of Washington is massive and concentrates on exploring misconceptions in physics in almost all content areas taught at the freshman level, and some taught in upper division courses. Throughout the analysis of laboratories any articles that the University of Washington has published referencing research on misconceptions relevant to that particular laboratory will be cited.

2.3.6 Current State of PER

The FCI and similar multiple choice examinations are the mechanisms that allow improved teaching techniques to be assessed and for this reason alone they are completely vital to PER and they were, in the early years, a unifying force that aided the field in developing legitimacy and in implementing educational reform.

The purpose of PER is not simply to catalogue the effectiveness of different educational techniques or to create a list of student misconceptions that only physics education researchers are aware of. The eventual purpose of the field is to reform the way that physics is taught to students in college classrooms and high school classrooms. Many of the active departments in physics education research are involved in the training of high school science teachers so the field in enjoying some success at this endeavor.

More problematic, perhaps, is the restructuring of the way physics is taught at the college levels. While several major universities have revamped their teaching curricula at the urging of major PER researchers in residence other schools are lagging behind in educational reform. The PER community has yet to develop a comprehensive strategy for appealing to the great majority of universities, who do not have physics education researches on staff who can direct and provide motivation for reform.

2.4 Laboratory History and Study

2.4.1 Laboratory Styles

The revised physics laboratory curricula use the problem based method of laboratory instruction, as opposed to the expository method previously used in lab courses. Techniques of laboratory instruction can be divide into four rough categories [34] as seen in

Table 2.3.

Style	Outcome	Approach	Procedure
Expository	Predetermined	Deductive	Given
Inquiry	Undetermined	Inductive	Student generated
Discovery	Predetermined	Inductive	Given
Problem-based	Predetermined	Deductive	Student Generated

Table 2.3: Descriptors of the Laboratory Instruction Styles

A call for laboratories to be administered by the inquiry method began to occur in the mid fifties [35], if not earlier. The introduction of laboratory work was always meant to stimulate students to recreate the experience and excitement of doing science; however, the techniques involved have always been subject to refinement. Henshaw's 1950 [35] article archives many personal statements by scientists concerning laboratory reform, and these are reflective of the prevailing climate of the times in the 1950s. In 1945 the Harvard University Press published *The Harvard Report, General Education in Free Society*, which sets forth a call for laboratory work based on the following principles:

> The scientist ensures himself, when he can, the proper circumstances for pursuing his inquiry by ordering the conditions of the natural events himself. This is the point of scientific experiment. By this means matters may be so arranged as to yield an unequivocal answer to a highly specific question concerning the real world. Such regulations of the system under regard is beyond the powers of students in other areas of reality. It is this constant appeal to things as they are which makes the direct experience of the field and laboratory essential in scientific education. Needless to say, this is only to the degree that work in the field or laboratory is designed not merely to keep students busy or develop technical proficiency, but to provide directly the materials of scientific argument and the tests of the scientific hypothesis. [36, p. 151]

The same publication speaks to methods of organizing laboratories in the following way:

> (The laboratory) should be planned to illustrate the methods by which phys-

> ical problems are solved. Every effort should be made to convey these as genuine experiences . . . in the scientific solution of material problems. [36, p. 228]

This is the philosophy that has most influenced recent laboratory reform in the physics department at the University of Arizona, as gleaned from conversations with the authors of the labs.

2.4.2 Historical Development of Physics Laboratories

Rosen [37] investigated the historical and cultural underpinnings that formed the impetus to incorporate laboratory instruction in classrooms throughout the United States. His article sets out the premise that the inclusion of laboratory instruction was motivated by two different philosophers, the first being Spencer (1820-1903) [38]. Spencer bases his argument for science being superior to all other forms of knowledge in that it has the greatest effectiveness for problem-solving in all imaginable contexts.

The other philosopher that Rosen cites as influencing academics to focus on teaching through experimentation was Goethe (1749-1832), whose work had a profound influence on German thought in the late 19th century [38]. Many American academicians attended graduate school in Germany and, upon their return to the United States, brought with them an appreciation for knowledge being developed from the perspective of the student. German scholars altered the application of Goethe's philosophy to be slightly more practical and materialistic, so the philosophy that began to dominate the German university system was that the specialist is the source of knowledge, and the specialist develops his skills through practical experimentation. This philosophy was fully expressed by 1910 with the *Arbeitsunterricht* (Constructive Activity) movement in which students were supposed to construct their own physical apparatus through the use of common laboratory tools, and thus establish a sense of one-ness with experimentation. Today, master machinists are still required to build their own equipment in order to earn certification, so the practical application of this philosophy has not entirely died out.

While the philosophical ground for physics being taught in an experimental setting was being laid, both secondary schools and colleges were slow to implement laboratory teaching, possibly due to the high cost of the purchase of instructional apparatus. The

Morill Act of 1862 set aside 30,000 acres of federal land for each legislator a state had in Congress for the purpose of establishing a college of mechanical and industrial arts. This strengthened the scientific infrastructure in the United Sates considerably and allowed for the development of technical colleges.

One of the fledgling universities founded around this time was Boston Tech, later called the Massachusetts Institute of Technology. When it opened its doors in 1869, it was the first college to require laboratory instruction [39]. One year after MIT's opening, the Harvard University Catalogue began advising students to pursue a course in mechanics with laboratory experimentation before application. This was followed by a series of stricter applications for entry, finally cumulating in the publication of a *Provisional List of Experiments in Elementary Physics*, which listed of recommended experiments. Once experimentation became a requirement for entering one of the United States' most prestigious universities, teaching laboratories were adopted in increasing numbers of secondary schools.

Harvard's list of experiments is recorded in Appendix B so that easy comparison can be made between the curricula outlined by Harvard in 1886 and the laboratory experiments commonly used in colleges and high schools today. The original list is quite extensive but was modified in 1887 to emphasize a course with a mixture of laboratory techniques, lecture, and the study of textbooks. The number of experiments was increased to sixty, of which only thirty-five had to be completed for admission to Harvard. The experiments are suggestive of a broad overview of many of the most important principles of physics, and most of the experiments described are still in use today.

> The two-decade interval, 1880-1900, saw the almost complete rejection of textbook work in high-school physics teaching as a method ”dead at the very root.” Instead, the modern laboratory became the classroom, because it did away with the reporter and brought the student face to face with nature. [37, p. 203]

The Harvard list is an impressive collection of experiments, and it is difficult to believe that a student who completed this list of experiments would not be eminently prepared for college; however, at this time laboratory instruction was implemented without systematic analysis of its educational benefit. Early questioning of the value of laboratory education occurred when the following remarks were made by Packard:

> In the laboratory, the student is introduced at once to the difficult subject of measurement, required to make immediate use of such unfamiliar instruments as the diagonal scales, the vernier caliper, and the balance sensitive to a centigram; to report his results in terms of the metric system, to discuss errors, sources of error, percentage of error, averages, and probabilities; to deduce laws, many of which he knew before, from data that cannot be made to prove anything, and to apply these laws to a set of problems that have no apparent relation to his immediate scientific environment, or to the questions that he is so anxious to have answered. [40, p. 881]

This criticism of the laboratory format stands largely unchanged to this day, however many scholars have made efforts to improve the quality of laboratory education.

2.4.3 Laboratory Reform

2.4.3.1 Call for Non-Cookbook Labs

Open-ended labs that are directed by TAs questioning their students may be implemented either with Socratic dialogue or through the use of leading questions. The difference between these two methods of approach is very important. In Socratic dialogue, as implemented by Socrates, an instructor responds to student statements by asking questions, sometimes without any clear goal at all. The idea is that students articulate their beliefs and, through articulation, engage in self-examination. Socratic dialogue leads naturally into inquiry-based labs, simply because the instructor does not necessarily have a target that the students are being led towards.

Leading questions, by contrast, are asked with a specific result in mind. When a lab instructor asks a leading question, it is designed to direct a student towards a particular idea or way of thinking about a problem. Leading questions tend to lead either into discovery or problem-based labs, since the labs have a definite goal, and the role of the instructor is to help the student move towards that goal with as little help as possible.

Prescott and Anger [41] were proponents of the idea of designing labs to be administered by leading questions. They described a laboratory in which all explicit instructions are removed, and replaced with questions designed to have students consider the design, execution, and interpretation of an experiment. In this format, the lab manual confronts students with questions such as “How many measurements should I make of each quantity?”, “What apparatus do I need?” and “What might I plot to get a straight line?”

Their laboratory curricula was shown to be very well received by the students through the use of an attitudinal survey; however, no effort was made to assess the learning gains achieved in such a course. Despite this, the authors expressed great confidence that their method of laboratory instruction was of great use to the students. They did note that the burden of implementing this style of laboratories falls upon the graduate student teaching assistants assigned to the lab, who require careful briefing on potential pitfalls in the labs and outlines of the sorts of guidance to be given to students, and are required to spend considerably more time preparing to administer the lab than they would with a conventional, or “cookbook”, lab.

2.4.3.2 Objectives of Laboratories

If laboratories are to be implemented, then what benefit shall they have for students? Two different opinions will be considered: first that of F. Reif and Mark St. John [42], and second that of Arnold Arons [43].

Rief and St. John consider that laboratories are valuable mainly because they can be used to aid students in the construction of basic and higher level skills. They consider that basic skills include:

1. Being able to use operational definitions to relate symbolic concepts to observable quantities. This skill subsumes the ability to estimate or measure important physical quantities at various levels of precision.

2. Being able to estimate the errors of quantities obtained from measurements. This skill involves applying habitually some qualitative or semiqualitative statistical notions, without any resort to excessive mathematical formalism

3. Knowing and applying some generally useful measuring techniques for improving reliability and precision. For example such techniques include making repeated

measurements, using independent measurement methods, or applying comparison or "null" methods. [42, p. 950]

The advanced skills include:

1. Being able to describe and talk about an experiment in a form easily understandable to other people (in particular, being able to summarize the most important ideas of an experiment and then to elaborate them to any desirable extent).
2. Being able to remember the central ideas of an experiment over a significantly long period of time.
3. Being flexibly able to modify the design or measurement procedures of an experiment when one is confronted with slightly different conditions or experimental goals. [42, p. 950]

Arons calls for laboratories to be structured in such a way so as to have students pursue one or more of the following modes of inquiry and critical thinking.

1. Observing phenomena qualitatively and interpreting observations.
2. Forming concepts as a result of observations.
3. Building and testing abstract models in the light of observation and concept formation.
4. By subjecting a piece of equipment to close examination in context, figuring out how it works and how it might be used (rather than simply being told how it works and what it is supposed to do).
5. Deciding what to do with a piece of equipment as well as deciding how many measurements to make and how to handle and present that data.
6. Asking or pursuing "How do we know...?, why do we believe...? What is the evidence for...?" questions inherently associated with a given experiment.
7. Explicitly discriminating between observation and inference in interpreting the results of experiments and observations.

8. Doing hypothetico-deductive reasoning (i.e., asking and addressing “what will happen if...?” questions) in connection with the laboratory situations. This includes visualizing, in the abstract, the effect of changing relevant variables or boundary conditions, visualizing outcomes in extreme or limiting conditions, and, where possible, forming an a priori hypothesis and then testing it by performing an appropriately designed experiment.[44, p. 334]

These lists contain similar language and highlight similar concepts, indicating a certain uniformity in the field of physics education as to what is expected from laboratory education. The lack of emphasis on uncertainty analysis in Aaron’s description of a highly functioning laboratory course is consistent with the passing of 20 years in educational thought. In general as time has passed, the focus of laboratories has shifted from teaching techniques of using laboratory equipment to students holistically interacting with equipment and forming their own ideas.

2.4.4 Methods of Laboratory Assessment

2.4.4.1 Performance Tests

Haym Kruglak published a series of articles comparing and contrasting different methods of assessing laboratory instruction. Prompted by discrepancies between student laboratory scores and scores on tests given in class, he developed and validated a method of testing students called performance tests [45], which involved students completing laboratory activities for a grade.

Kruglak and Trainor [45] developed approximately 600 different laboratory performance tests, all of them short experiments that students can perform at a test station within a few minutes so a group of students can complete several different test stations during the examination time. After these items were constructed, they were sent to four different college teachers who rated them independently, and then the highest scoring questions were tested over the course of two years in every physics lab at the University of Minnesota. Statistical analysis revealed that instructors who were provided with detailed grading keys were able to grade essentially identically. Student grades on performance tests correlated slightly with theoretical tests given in the associated lecture. Kruglak felt this correlation was due to overlapping of vocabulary and concepts between experimental and theoretical physics and he proposed that performance tests be used as a method of laboratory assessment in order to probe students' experimental skills.

Kruglak designed both short and long performance tests. The following are examples provided in Kruglak's 1951 [46] article of the short-item and long-item performance tests, respectively:

> *Given:* Meter stick supported at the center of gravity, weights, sliding hooks, unknown weight *Problem:* To find the unknown weight *Note:* Neglect the weight of the sliders. When instructor signals, move to location No. 10. [46, p. 956]

> *Given:* Derrick, weight, spring balance, and meter stick *Problem:* Find the weight of the boom *Note:* The mass of the suspended weight is marked on it. When instructor signals, move to location No. 25. [46, p. 956]

2.4.4.2 Paper-pencil tests

In addition to his work on performance tests, Kruglak compared and contrasted student achievement on performance tests with their achievement on written tests in a few different contexts. In two different articles [47], [48], Kruglak explored the relationship between written tests and performance tests. In these investigations he tried to design tests that would, without recourse to students using equipment, evaluate such things as:

1. identification of apparatus from photographs, diagrams, and descriptions.
2. knowledge of the function of apparatus, materials, and instruments.
3. reading instruments commonly found in the physics laboratory.
4. symbolic representation of real objects and identification of objects from their symbolic representation.
5. selection of apparatus for specific measurement or function.
6. perform appropriate calculations from specific data.
7. reading and interpretation of graphs.

8. drawing a graph from a set of experimental data.
9. understanding directions for carrying out a specific laboratory procedure.
10. recognition of random and systematic errors.
11. conducting an error analysis from experimental data.

Various written multiple-choice laboratory tests were constructed. A year-long study yielded interesting results. The paper-and-pencil tests administered as pre-tests were good predictors of post-test scores on other paper tests [48], so Kruglak concluded that laboratories are effective at making good laboratory students better, but did not have a strong effect on those entering with poor laboratory skills. Another interesting result of the study was that female students scored substantially lower than male students on the questions dealing with identifying equipment and labelling the function of equipment used in the mechanics labs.

Further studies [45] tested an essay-response question type, a multiple-choice paper-and-pencil analog of the performance test, and the traditional performance test. Analysis of the correlations among the scores on these different items revealed that there was a low correlation between all performance tests and the analogous items on the multiple choice and essay type paper tests. This result led Kruglak to select performance tests over written examinations as a method of effectively measuring the learning of experimental physics. Kruglak, confident that he had validated his performance tests as a method of assessing the skill set that students developed in the laboratory, eventually distributed them to over 1,000 colleges with the aid of an NSF grant [49]. This distribution of his hands-on tests helped to shape the method by which laboratories are assessed throughout the country.

2.4.5 Analysis of Laboratory Effectiveness

2.4.5.1 The Experiments of Hans Kruglak

Systematic laboratory assessment in the United States did not begin in earnest until approximately 1950 when Haym Kruglak conducted some precisely-controlled but low-sample-size experiments at the University of Minnesota to examine populations of students. In the first of these experiments, populations of students taking mechanics who performed laboratory work from a lab manual were compared to groups of students who had experiments demonstrated to them [50]. In this study Kruglak separated forty-six students into two control groups, who took labs, and two experimental groups, who watched demonstrations and were not allowed to manipulate the equipment until the final exam. Attrition and other factors limited the final study to thirty eight students. He introduced four different methods of testing learning gains in the mechanics laboratory. He tested students with short-item laboratory performance tests, long-item laboratory performance tests, a written laboratory test, and a mechanics theory test.

Kruglak's analysis of his results revealed that the only statistically significant difference between the performance of students who watched demonstrations and the performance of students who performed laboratory experiments occurred on the short-item performance tests. No statistically significant differences were noted for students with different laboratory professors, nor were any differences traced to different laboratory techniques. This result is interesting as it speaks to the equality of effectiveness of laboratory demonstrations and manipulation of laboratory equipment as learning tools. This experiment does not, however, make any firm conclusions concerning the value of either technique versus having no experimental instruction whatsoever.

Kruglak's second extensive laboratory experiment [51] compared and contrasted students in a mechanics class who either experienced laboratory by demonstrations or by doing individual work, and students who did neither. He again assessed students using written tests focusing on theoretical understanding, short-item laboratory performance, and long-item laboratory performance. This time, the students who took the hands on laboratory and the students who watched the demonstrations in laboratory both performed better than the no-laboratory control group on the laboratory written exam, the short-item laboratory performance, and the long-item laboratory performance tests. This difference was statistically significant; however, no statistically significant difference was found for performance on the mechanics theory tests. Thus Kruglak concluded that laboratory instruction (either through demonstration or through manipulation of equipment) does not significantly influence performance on theoretical paper-pencil tests, but does influence performance on tests designed to measure laboratory outcomes.

2.4.5.2 The Experiments of Sanborn Brown

One of the most interesting studies focusing on laboratories was undertaken in the 1950s by Sanborn Brown, who designed an experiment that tested the effectiveness of laboratory instruction in high school [52]. Brown had just finished two survey articles, one analyzing laboratory techniques used in England [53], and another analyzing the laboratory techniques used at MIT [54]. Having established his familiarity with college instruction, he now focused on evaluating the effectiveness of physics lab instruction in high schools. Brown's experimental design was to survey all high school and preparatory schools that had sent students to MIT, learn what their laboratory curricula were, and obtain copies of their laboratory books (if available). Brown selected six experiments that were essentially equivalent to those performed in high school, and asked 900 freshmen, of which between 300-400 had completed some of the experiments before, to answer questions on the experiments. Brown's found no statistically significant differences based on analysis of students' laboratory grades, indicating that students were not retaining information from laboratories that they had been exposed to in high school.

Brown then instituted another experiment to test whether the students could recognize the apparatus used in high school by handing the students pictures and asking for the name of the apparatus, what physical quantity was measured with the apparatus, and a description of how the experiment was performed. The result of Brown's study was that only 41% of the students were able to correctly recognize apparatus that they had used in a high school class.

Brown refined his study to look for further correlations, and separated the population into groups that had taken physics labs as high school seniors and as high school juniors. He discovered no statistically significant differences between the populations based on their ability to identify equipment related to laboratories that they had taken. Brown also gave the test on identification and use of laboratory apparatus to 84 freshmen who had never taken laboratory physics at all, and found that they averaged significantly lower on the exams (30% as opposed to 41%), indicating some advantage to having had a physics laboratory in terms of developing a skill set necessary to pass exams on laboratory theory.

Various other statistical studies were used by Brown to explore his data set, but he eventually concluded that he could find no obvious benefit to high school laboratory instruction beyond student testimonials that laboratory experiences in their secondary education helped convince them to choose science or engineering as a profession.

2.4.5.3 Miscellaneous Experiments

In 1985, Long published a study in which he investigated the influence of physics laboratories on performance in the attendant lecture course [55]. Long's study compares a population of 2449 students in physics courses, of which 70% took a laboratory with the lecture course. He is careful to note that different majors are required to take the laboratory courses and hence the distribution of students taking the laboratory course is not random.

However, his statistical results are quite revealing, and similar to Brown's result with a more stratified analysis of population by grade in the lecture course. Long found that students whose grades in the lecture course were 3.0 (A) or above seemed to have a very slight positive increase in their score as a result of taking the laboratory. Students with grades below 2.0 (C) seemed to suffer from taking the laboratory at the same time as the course, and those whose grades were between a 2.0 and a 3.0 (B) saw a grade increase of approximately a third of a point (from a B- to a B, for example). Thus Long's study bore out Brown's result that laboratory learning does not significantly impact theoretical understanding of physics as measured by paper-pencil tests.

While some publications disclaim the tangible intellectual gains (as measured by paper and pencil tests) to be had from laboratory work, other publications explore the positive impact of laboratories. For example Shonle published an article [56] in which he reports students' reactions to open-ended laboratories. He had overseen the implementation of open-ended physics labs at the University of Colorado and, as part of the assessment of his course, he had distributed a Likert-type scale showing extremely positive student response to the lab format in terms of students' self-assessment of the quality of their learning and enjoyment of the laboratory. The statistical results of his survey are reproduced in Table 2.4.

So student reaction to laboratory work, especially open-ended laboratory work, can be very positive. It is worth noting that Shonle's vision for open-ended labs stressed working

Statement	Strong Agreement	Agreement	Neutral	Disagreement	Strong Disagreement
I learned something in this laboratory	30	34	3	0	0
I felt that this laboratory is better than one using the conventional format	39	21	4	0	0
I had fun	32	31	4	0	0
This laboratory was valuable	25	31	8	1	1
There should be no laboratory at all	0	1	10	14	41

Table 2.4: Student Reaction to Open Ended Laboratories as Investigated by John Shonle

on the same project for multiple weeks, and he also emphasized a form of scaffolding where students received a significant amount of help on the first lab and successively less help on later labs.

Another researcher whose work reflected positive gains to be had from laboratory achievement was Michael Freedman [57], who instituted a study on the relationship between laboratory instruction and attitude towards and achievement in science. He found that 9th grade inner-city science students spread across 20 different classes experienced increases in positive attitudes towards science and performance on in class tests due to their laboratory participation. This result is not directly relevant to college laboratory work, but is one of a handful of studies citing laboratory work as having value for students. Overall, the research on laboratory education appears to indicate that it has value within the context of secondary schools during the time that students are in school, but promotes little long-term learning as measured by student retention when they enter college.

2.4.5.4 A Rejection of Laboratories

W.S. Toothacker writes “Is student laboratory work worth this great expense of time and money? The answer, based upon studies which will be discussed below, appears to be no” near the beginning of his 1983 review article [58, p. 517]. Toothacker proceeds to review and discuss many of the articles discussed in this chapter. His recommendations to the physics community is twofold. Firstly, he concluded that the body of evidence indicated that students taking freshman and sophomore labs…

> "…are not learning laboratory techniques, methods, and familiarity with equipment from their laboratory work and the majority of them do not need to as they will never work in a physics laboratory. Thus all freshman and sophomore physics laboratory classes should be abolished. Laboratory work in physics should be reserved for those students who need to learn how to do experimental physics, i.e., those students seeking a degree in physics." [58, p. 519]

Toothacker's second recommendation is that, for those students who require it, laboratory work be replaced with a course on experimentation. This course could be uncoupled from a lecture course and taught as a separate subject in which topics such as "data analysis, experimental graph plotting, curve fitting, accuracy, precision, significant digits, estimation and propagation of uncertainties, the difference between random and systematic errors, and similar topics" [58, p. 519] are taught. He notes in his paper that such a course was designed and executed at the University of Liverpool with what were judged by the faculty at that institution to be good results.

2.4.6 Gender Studies in Laboratory and Science Education

The body of literature developed on the study of differences in science achievement between males and females is quite extensive. There is some controversy about the nature of performance differences between the two genders, with some articles [59] noting relatively little difference in manipulative skills demonstrated by males and females and others [60] noting significant differences in cognitive abilities as measured by performance tests. So, while some controversy exists about the relative level of abilities of female and male students, it is however undoubtedly true that females are, as of 2006, less common in the physical sciences at the college level than males, and much research has been done to probe why this may be the case.

Weinburgh conducted a meta-analysis [61] of all available studies on attitude towards science between the years of 1971-1990 in the journals *Science Education*, *School Science* *and Mathematics*, *Journal of Research in Science Teaching*, and *School Science Review*. During the research period there were 18 usable articles and 31 different populations to be studied. For general science, the correlation between attitude and achievement was .5 for females and .55 for males, so a significant relationship between attitude towards science and performance in classrooms was found. Eighty one percent of the studies surveyed reflected more positive attitudes towards science in males than in females.

The part of the study that dealt specifically with physics was much more modest, including only two studies. The correlation between attitude and achievement for physics is much lower than for science as a whole (males = .34, females = .37) indicating that many high achievers in physics do not have a positive attitude toward science. The article notes that females in physics classes past the 10th grade had approximately the same attitude towards physics as their male peers.

Some well-reasoned suggestions have been made about methods of improving attitudes towards physical science among females. Linn, de Benedicitis, DeLucchi, Harris, and Stage [62] note in their 1987 article that 17 year old females are more likely to use the "I don't know" response than their male peers when asked questions with physical science content or questions centered around traditionally masculine themed activities, such as football. This lack of engagement with the topic is fatal to inquiry-themed methods of instruction, in which student engagement and curiosity are critical. The authors advise that science courses be revised to omit items with either traditionally masculine content or esoteric content knowledge so that the courses are overall more comprehensible and approachable.

Freedman [63] has also done studies on the effects that laboratory instruction can have on male and female students' attitudes toward science. His article begins by noting through citation and by direct assertion that females are much less likely to handle laboratory equipment than males. Freedman thus designed a study in which ninth grade physical science student were divided into two groups and one group was exposed to no laboratory work at all, and the other group was exposed to weekly laboratory investigations in contexts that did not allow females to opt out of the experiments. During the course of this study females in the experimental group were supervised to ensure that they had equal amounts of time manipulating laboratory equipment as their male counterparts.

His study found that, in the context of a 9th grade science course, exposure to laboratory work significantly improved attitude towards science and achievement in science for the female experimental group in comparison to the female control group. Boys benefitted from the laboratory instruction as well, but not to the same extent. This result is significant insofar as it helps to establish the cultural background of females entering the university physics program and their potential laboratory background. It is quite possible that many of them experienced a high school culture which allowed them to opt-out of laboratories as described in this article.

2.5 Miscellaneous Research Summary

2.5.1 Active Engagement Techniques

Active engagement, or interactive engagement, is a catch phrase for a body of techniques that promote inquiry-based learning, in which students construct their own knowledge via discussion with each other, interaction with equipment, or other such techniques. Inquiry-based learning, in which students are responsible for posing and answering their own questions with guidance, has been shown to be much more effective than traditional lecture or laboratory manipulation in achieving student understanding [64], so the promotion of active learning is a prominent goal of PER. The body of literature supporting methods of implementing active engagement techniques is large, however much of it deals with implementing these techniques in lecture halls [65] rather than laboratory classrooms; thus, only the literature specifically dealing with forming groups, posing appropriate questions, and managing student discussion will be discussed here.

Very little research has been done on active engagement techniques that are specifically for usage in laboratory contexts, hence the analysis of well established and validated techniques of promoting student engagement is valuable insofar as it can inform techniques of interactive engagement in the laboratory. The active engagement techniques that were eventually implemented in the second semester lab after it was revised were not as intricate as those described here due to concerns about the teaching assistants' ability to effectively deploy such techniques without additional training.

2.5.1.1 Cooperative Grouping

Patricia Heller and her collaborators [66] [67], have had much to say on the topic of cooperative learning. She implemented group work on physics problems at the University of Minnesota in the early 1990s and published a study of the effectiveness of such group work. The group size was three students, and the three group members alternated between roles of Manager, Skeptic, and Checker/Recorder. She found that structuring groups this way was effective in mitigating gender issues within groups and in helping to keep all students involved regardless of their relative skill levels in physics. Although this method of administering group work was not used in the revised laboratories it is included here as an illustration of techniques that were considered and then rejected as being unduly complicated for TAs to implement without extensive training. Ideally group work will be a highly structured until students learn good group work habits.

The class began with the explicit teaching of appropriate problem-solving strategies for group work. All the students in the class were taught to solve problems by explicitly going through the following steps, and were monitored to see how the students interpreted and executed these instructions. These steps are described in Heller's own words:

1. **Visualize the problem:** This step is a translation of the problem statement into a visual and verbal understanding of the problem situation.

2. **Physics description:** This step requires students to use their qualitative understanding of physics concepts and principles to analyze and represent the problem in physics terms (e.g. vector diagrams).

3. **Plan a solution:** This step involves translating the physics description into an appropriate mathematical representation of the problem (equations of the principles and constraints) determining if there is enough information represented to solve the problem, then specifying the algebraic procedure to extract the unknown variable(s).

4. **Execute the plan:** Students use mathematical rules to obtain an expression with the desired unknown variable on one side of the equation and all the known variables on the other side. Specific values are then substituted into the expression to obtain a numerical solution.

5. **Check and evaluate:** Finally, the students evaluate the reasonableness of their answer — are the signs and units correct and does the answer match their experience of the world and/or their expectations of how large the numerical answer should be?

   [66, p. 629]

Heller stressed the importance of choosing appropriately hard questions for group work, and provided a guide to designing problems of appropriate difficulty level. Problems should always be conceptual in nature and should be evaluated in terms of six characteristic indicators to determine the difficulty of the problem. Again in Heller's words, the factors that add to the difficulty of problems are:

1. **Problem context:** Contexts familiar to the majority of introductory students through direct experience, newspapers, television, or solving standard textbook problems are easier than problems with contexts unfamiliar to the students (e.g., ion beam, protons from the Sun, x-ray signals from pulsars).

2. **Problem cues:** Problems that cue a standard application of a set of related principles to solve the problem are easier than problems that do not explicitly cue a standard approach. For example, force problems that specify a force as the unknown variable (e.g., what is the lift on an airplane?) are easier than force problems that specify a mass as the unknown variable (e.g., What is the mass of the planet?).

3. **Given information:** Problems with no extraneous information or missing information in the problem statement are easier than problems with irrelevant information or missing information that must be recalled or estimated.

4. **Explicitness of question:** Problems that specify a particular unknown variable (e.g., What is the muzzle velocity of the bullet?) are easier than problems for which the desired unknown variable must be determined (e.g., Will this design for the lunar lander work?).

5. **Number of approaches:** Problems that can be solved with one set of related principles (e.g. kinematics or energy conservation) are easier than problems that

require the application of more than one set of related principles for a solution (e.g., both kinematics and energy conservation).

6. **Memory load:** Problems that require the solution of five or less equations are easier than problems that require the solution of more than five equations. [66, p. 631],

Heller rating each of the six categories with a boolean grading system of 0 (easier) or 1 (more difficult) and assigned the difficulty of a problem to be the grades of all six categories added together. He confirmed that students had lower performance scores on higher difficulty problems. This suggests that problems given in discussion sections should be carefully designed and tested to stress students' reasoning abilities, but not to be so difficult that a group of two or three cannot solve them. Ideally, the level of difficulty of discussion questions should be increased as the semester proceeds, but should seldom pass a difficulty rating of 3 or 4.

2.5.1.2 Think-Pair-Share

In Mazur's 1997 book *Peer Instruction* [68] he describes a method of active engagement that can be used in a classroom called Think-Pair-Share. The steps of Think-Pair-Share are exactly as follows:

1. Before class begins, select a question that fulfills the following criteria:
    (a) The question focuses on a single concept.
    (b) It is not solvable by relying on equations.
    (c) It has adequate multiple choice answers.
    (d) It is unambiguously worded.
    (e) It is neither too easy nor too difficult.
2. Pose the question in class (take no more than one minute).
3. Give students time to think silently about question (one minute).
4. Students record individual answers or (optional) use clickers to transmit answers en masse to instructor.

5. Students convince their neighbors of the correctness of their answers (1-2 minutes).
6. Students record new answers or use (optional) clickers to transmit answers en masse to instructor.
7. Instructor or selected student explains correct answer (2 + minutes).

Students think individually about the correct answer to a question, pair with their neighbors, and then try to convince them of the correctness of their answer. Think-pair-share is valuable both as a tool of forcing students to articulate their beliefs about physics in a non-threatening environment, and as a method of encouraging discussion about physics. An example of an appropriate think-pair-share question is taken from *Peer Instruction.*

> Consider a capacitor made of two parallel metallic plates separated by a distance d. The top plate has a surface charge density $+\sigma$, the bottom plate $-\sigma$. A slab of metal of thickness $l < d$ is inserted between the plates, not connected to either one. Upon insertion of the metal slab the potential difference between the plates
>
> 1. increases
> 2. decreases
> 3. remains the same [68, p. 194]

Such questions allow faculty members to assess whether students have a good conceptual understanding of physics, while allowing the students themselves to examine the extent of their knowledge in a non-threatening discussion with a peer, as opposed to an authority figure. Think-Pair-Share is probably the most effective active engagement technique that can be used in the large lecture environment, but it could be highly valuable for probing student understanding at various stages of a laboratory if TAs were provided with think-pair-share questions ahead of time and received adequate training in how to administer such questions.

2.5.2 Wait Time

Wait time is the name given to the amount of time an instructor pauses after a question is asked (wait time type 1) and after a question is answered (wait time type 2). A more exact description of the different types of wait time is [69] given below using game theory to explain the interacting, where a teacher counts as one player and all of the students in the classroom count as another player.

1. *"Wait-time of the species one type"* may appear in two varieties. Normally it begins when the teacher stops speaking and terminates when a student responds or the teacher speaks again. If, as sometimes happens, a teacher asks a question, pauses, calls on a student, and pauses again, the two pauses are summed. Together they constitute an instance of the first species of wait-time.
2. *"Wait-item of the species two variety"* is calculated by taking the sum of all pauses occurring on the student player side and terminates when the teacher speaks. The more common varieties include the case where a student speaks, stops, and the teacher speaks again; and a student speaks, pauses, speaks again, and the teacher rejoins the play. (Note: Here the term student is used generically, i.e., refers to the two-player model. These pauses may occur within the speech of a single pupil or they may occur between the speech of a succession of pupils. In either case, the collection of pauses are summed and constitute a single instance of the post-student response wait-time.) Correlations between the two species of wait time vary somewhat but tend to be on the order of 0.17.

Most of the research done on wait time was conducted by Rowe over the course of several decades [69], [70]. The body of her research centers on the effect of increasing wait time on primary and secondary school students, however analysis of college and high school seminar tapes reveals that the average wait-time in these environments ranges from 1 second to 2.8 seconds with a mean of 1.8 seconds.

In 1992 Duell published a study [71] in which she determined that the mean wait time evidenced by education professors was 1.89 seconds for questions requiring students to recall facts and varied between 1.48 and 2.94 seconds for higher cognitive level questions. Since this was a study of professors in education who are aware of the existence of the concept of wait time and may have made efforts to implement the technique it seems reasonable to assume that professors in physics and graduate students in physics who have no knowledge of wait time at all would evidence similar or lesser wait times.

Duell continued her study in 1994 [72] when she designed an experiment to measure the effect of using 3 seconds or six seconds worth of type 1 wait time on a group of education undergraduates. Again, the population may not be generalizable to physics students, however this is the only serious experiment comparing and contrasting the effects of different amounts of wait time on college students.

In her first experiment, Duell took 63 education students and divided them into four different experimental sections, each of size 13-18. Two of the sections were administered with wait times of three seconds, and two were administered with wait times of six seconds. The lectures, questions, and demonstration materials used in the classrooms were all prescripted, and the instructors used watches with second hands to try and match the wait time requirements as closely as possible. The groups assigned to three seconds received 2.80 ± .87 seconds wait time for each question and the groups assigned to six seconds received 6.37 ± .91 seconds for each question.

The effectiveness of instruction using the different amounts of wait time was assessed with a test that was divided between 11 low-level knowledge items and 19 higher-level knowledge and synthesis questions. No significant differences were noted between the three second and six second groups in scores on the low-level knowledge items. However a difference was found between the three second and six second group scores on the synthesis questions; the students who were treated with six seconds worth of wait time underperformed the three second group by a statistically significant amount.

Analysis of recordings in the classroom suggests that the students were aware of the remarkably long pauses and spent time discussing them. She next designed a new experiment, nearly identical in design to the previous one, to assess the difference between one second of wait time and three seconds of wait time. A total of one hundred and thirty one students participated and the prescripted classroom materials and other details of the study were not changed. This time no statistical difference was found between the groups on either the low-level knowledge and the high-level knowledge/synthesis items on the test.

These experiments appear to make a compelling argument against the use of six second wait-time in college classrooms, however the method used to assess the effects of wait time, that of a pen and paper test, may not be the most effective way to explore the full effects of wait time. For example when Rowe was studying the effects of wait time on children she found that the introduction of wait time caused the following changes in behavior:

1. The length of response increases.
2. The number of unsolicited but appropriate responses increases.
3. Failures to respond decrease.
4. Confidence as reflected in decrease of inflected responses increases.
5. Incidence of speculative responses increases.
6. Incidence of child-child comparisons of data increases.
7. Incidence of evidence-inference statements increases.
8. The frequency of student questions increases.
9. Incidence of responses from students rated by teachers as relatively slow increases.
10. The variety of type moves made by students increases.

Duell's study, however well designed, was rather limited in the techniques it used to measure the effect of wait time on college students. Kenneth G. Tobin [73] conducted a study analyzing the dependencies of some of these variables on the use of wait time. In his study wait time was controlled in 23 Australian middle school science courses during identical science lessons. The effect of the alteration of wait-time was assessed using a series of multiple-choice examinations.

Thirteen teachers were assigned to use a wait time of three seconds, of which only eight actually succeeded in waiting the full three seconds. The students in this classroom outperformed the other groups, which recieved lesser wait times, on the written tests. The experimental design of this study was similar to Duell's, however Tobin found a statistically significant gain between the courses using extended wait time and the rest of his sample. The difference in the results of the studies could be due the difference between middle school students and college students, the difference between science courses and education courses, or the difference between the way that teaching is organized in middle school and the way that it's organized in college.

2.5.3 Qualitative Research in Social Sciences

Before any research project involving social sciences is begun, questions such as "What is the bias of the social researcher?" "What is the goal of the project?" and "How will project evaluation be carried out?" need to be answered adequately. In this section a brief overview of perspectives and techniques from the social sciences will be described so as to acquaint the audience with considerations relevant to carrying out an investigation. In the techniques portion of the dissertation, the perspectives chosen and techniques utilized in this particular project will be explored.

2.5.3.1 Perspectives on Research in Social Sciences

Schram [74] makes a strong argument that all social sciences researchers, even those who do qualitative research, approach their research with a mixture of biases and intentions that lead to their looking at their research results from different perspectives, which Schram calls "lenses". Schram describes four of these different lenses and suggests that researches should consider carefully which of these lenses is used to describe their research, or if their research is a synthesis of two or more lenses.

The Interpretive Lens

Schram asserts that social scientists who do research through an interpretive lens believe that knowledge is constructed entirely through people interacting with each other in different social settings, thus the aim of social research is to understand the subject's environmentally-dependent point of view. A researcher working through an interpretive lens is required to actively engage their research subjects so as to synthesize and interpret their viewpoints and to understand the constructs that the subjects design to deal with their environment.

The Critical Lens

Those social researchers who examine their subjects through a critical lens work with the expectation that their work will be used to bring about change in the environment that they are studying, thus their research embraces the twin concepts of advocacy and activism.

The Feminist Lens

Social scientists who study their subjects through the feminist lens are putting their focus and attention on oppressive social structures that exist in the object of their study and the means to challenge and change them. Particular attention is placed on barriers that may exist from the consideration of race, class, culture, ethnicity, gender, or other divisions of identity.

The Ecological Lens

Researchers that work through the ecological lens think of humans as being embedded in a social context that influences their behavior, thus the social researcher concentrates on understanding the social dynamics of the group that is being studied.

The research perspective used in this dissertation is a combination of the feminist lens and the critical lens. The labs were studied both with an eye to what can be revised to make them more effective, and with critical attention to the ways in which the implementation of the labs and the writing in the labs themselves could possibly marginalize female students or ethnic minorities.

2.5.3.2 Criteria for Design of Observation Protocols

In order to avoid observer bias, this study was designed to be as quantitative as possible. Most of the techniques developed for use in this dissertation were taken from the book *Qualitative Data Analysis* by Miles and Huberman [75]. It contains detailed descriptions about methods to design useful qualitative and quantitative research which were used to inform how the instruments used in this dissertation were designed.

Qualitative research is conducted through the use of tools called instruments, which are coded protocols used to record different types of events. Most instruments used in this study were developed within the first week or two of observations each semester. Miles and Huberman present the following justifications for developing instrumentation after the study has begun:

1. Predesigned and structured instruments blind the research to the site. If the most important phenomena or underlying constructs at work in the field are not in the instruments, they will be overlooked or misrepresented.
2. Prior instrumentation is usually context-stripped; it lusts for universality, uniformity, and comparability. Qualitative research is the one place where contexts can and should be studied; it is the particularities that produce the generalities, not the reverse.
3. Most qualitative studies involve single cases, with few people involved. Who needs questionnaires, observation schedules, or tests-whose prime function is to yield economical, comparable, and parametric distributions for large samples?
4. The lion's share of fieldwork consists of taking notes, recording things (conversations, meetings), and picking up samples (documents, products, artifacts). "Instrumentation" is a misnomer. Some orienting questions, some headings for observations, a rough and ready document analysis form are all one needs at the start-perhaps all one will *ever* need in the course of the study. ([75], p 42)

Coding with Instruments

It is vitally important to any field researcher to develop a series of codes to organize and record information. They allow for quick recording of information, and allow a researcher to condense and quickly collate information. Miles and Huberman offer the following advice on construction of codes:

- Make certain the all the codes fit into a structure that they *relate to* and are *distinct from* others in meaningful, study important ways. Don't casually add, remove, or reconfigure codes.
- Keep the codes semantically close to the terms they represent. Don't use numbers as codes.
- Have the codes on a single sheet for easy reference.
- Define the codes operationally and be sure all analysts understand the definitions and are able to identify rapidly a segment fitting the definition. [75, p. 64]

The above criteria were used to create all codes used in this project. During the project, various types of codes were developed to describe potentially relevant behaviors on the part of the TAs or the students in the laboratories.

Marginal Notes and Reflective Notes

While field observations are in progress, events sometimes occur that are worthy of note. Marginal remarks are interspersed within the codes in order to keep notes about what is happening in the environment and to add greater clarity to the field notes. Reflective notes are used to interpret field notes, and should be written when field notes are reviewed and before they are used for analysis. Examples of the use of reflective notes include:

- what the relationship with the respondents was like
- second thoughts on the meaning of what a respondent was saying
- doubts about the quality of the data being recorded
- a new hypothesis explaining what was happening
- a mental note to pursue an issue further in the next contact
- cross-allusions to something in another part of the data
- the researcher's own feelings about what was said or done
- elaboration or clarification of something said or done [75, p. 64-64]

Document Summary Form

Document summary forms are created to put field notes and codes in context. They are written so that they all follow the same format, and summarize a distinct period of observation or category of observation. In the context of this dissertation, the notes on each individual lab comprise a document summary form.

2.5.3.3 Interview Background

Interviews conducted in this study are semi-structured, as defined by Denscombe:

> With semi-structured interviews, the interviewer still has a clear list of issues to be addressed and questions to be answered. However, with the semi-structured interview the interviewer is prepared to be flexible in terms of the

> order in which the topics are considered, and, perhaps significantly, to let the interviewee develop ideas and speak more widely on the issues raised by the researcher. The answers are open-ended and there is more emphasis on the interviewee elaborating points of interest. [76, p. 167]

Since all interviews in this study were conducted on a volunteer basis, there was little control over the number of students who were to be interviewed and the number of students who actually appeared at the scheduled times. Consequently, some of the interviews were one-on-one, and some were group interviews with three or more people. The procedure for interviews changed depending on the number of people participating, as the techniques of managing individual interviews differ from group interviews.

One-on-one interviews are valuable, as the opinions and insights of one person are recorded and explored thoroughly. They are both easier to control than group interviews and easier to keep on topic. They also have the great advantage of providing an environment in which shy subjects are able to take all of the time that they need to articulate their thoughts without pressure from a group.

Group interviews, on the other hand, are used to stimulate discussion among the participants in order to try and reveal any consensus views that may exist among the participants. Careful efforts must be made to make sure that quieter participants have an opportunity to voice their opinions, and to make sure that all participants feel free to voice all their viewpoints, including those that are not consistent with whatever consensus view may be held by the rest of the group.

CHAPTER 3

TECHNIQUES

This chapter describes the techniques used to perform classroom observations, the methods used to design the weekly quizzes, the techniques used to prepare TAs for classes, and the interview protocols used. Thus, it is a description of the qualitative methods used to assess the effectiveness of the revised labs and the qualitative methods used to improve laboratory performance in 2005 during the revision cycle of the dissertation.

Key research questions that guided this study were:

1. What are the behaviors that are induced in the TA by the setup of the TA manual and the lab meetings?
2. What is the impact of the rewritten laboratories on TA behaviors and implications for future TA training?
3. What knowledge and skill sets are acquired by students as a result of exposure to the laboratory curricula?

3.1 Observation Protocols

3.1.1 SATIC

The instrument *SATIC* was developed by Abraham and Schlitt [77] for self-criticism of teachers. Ideally, teachers record themselves, make a transcript of their conversation, and sort each "comment" that they make into a different category. A single comment is defined by the uninterrupted period of time that A teacher is speaking. The fourteen SATIC categories are:

1. Lectures or gives directions (INITIATORY BEHAVIOR)
2. Makes statement or asks rhetorical question (INITIATORY BEHAVIOR)
3. Asks short answer question (INITIATORY BEHAVIOR)
4. Asks extended answer question (INITIATORY BEHAVIOR)

5. Pauses (gives enough time for student response) (INTERIM BEHAVIOR)
6. Cuts off or rushes student response (INTERIM BEHAVIOR)
7. Rejects student comment, answer, or question (RESPONDING BEHAVIOR)
8. Accepts student comment or answer (RESPONDING BEHAVIOR)
9. Confirms student comment or answer (RESPONDING BEHAVIOR)
10. Clarifies or interprets student verbal behavior (RESPONDING BEHAVIOR)
11. Answers student question (RESPONDING BEHAVIOR)
12. Asks student to clarify or elaborate (RESPONDING BEHAVIOR)
13. Uses student question or idea (RESPONDING BEHAVIOR) [77, p. 677]

The SATIC format was used for the first semester of observations, however it became clear that using this protocol to analyze every lab session would be prohibitively time-consuming. After consideration of what the SATIC was designed to measure and what was important to analyze in the classroom, the number of categories was reduced from thirteen to four:

1. Initiatory (Talking): In this type of behavior the instructor either lectures or asks rhetorical questions.
2. Initiatory (Questioning): Here the instructor questions the students.
3. Responding (Teacher-Centered): Here the instructor responds to the student in such a way as to not require students to respond further.
4. Responding (Student-Centered) : Here the instructor answers a question in such a way as to require the student to still participate, for example by asking a student to elaborate on his or her original question. According to traditional SATIC format, this category also includes situations where the instructor incorporates a students' statement into his or her own lecture.

Because TAs sometimes speak for long periods of time without any student input, TA behavior is coded into a separate category for every fifteen seconds that the lecture lasts, rather than being marked separately for every behavior between student interactions. The fifteen-second observation period made observations of interim behavior impractical, as this would occur on a timescale that the observation is insensitive to. The new categories were contrived to represent the remaining 11 SATIC behaviors as accurately as possible.

Thus, in the first semester of this study the observer sat with a stopwatch, and watched the lecture. Every fifteen seconds during the lecture the observer binned the TA's comments into one of the four modified SATIC categories.

3.1.2 ROP Description

The following is a description of the Revised Observation Protocol (ROP) as applied to observations of TAs. This protocol was designed by the author of this dissertation and implemented to replace the revised SATIC in the Fall of 2004. The ROP has several different categories designated by capital letters, all of which can be inserted into either student-centered or teacher-centered activities.

- Teacher-centered
  - **T** The TA lectures, gives directions, or asks rhetorical questions.
  - **B** The TA writes on the blackboard as part of her or his lecture (direct instruction with enhancement activity).
  - **E** The TA manipulates equipment as part of his or her lecture (direct instruction with enhancement activity).
- Student-centered
  - **Q** The TA asks a specific student or the class as a whole a question and waits for a response.
  - **A** The TA is stopped in his or her lecture by a student question, and responds to the question before continuing.
  - **W** The TA requires a student to present work on the main whiteboard or on the student's whiteboard.

While the ROP observation is being conducted, the TA is analyzed in fifteen second increments. Every behavior the TA exhibits for these fifteen seconds is noted and recorded, and then the box representative of that time slot is scored according to the following criteria: if Q, A, W, B, or E are observed to occur then they must be recorded. If more than one of Q, A, W, B, or E is noted as occurring in a single time block, then the time block is recorded as one of these behaviors and the remaining behaviors are moved to blocks of time previously scored as T. In this way, Q, A, W, B, and E are all maximally expressed and T is always undercounted.

This observation protocol succeeds primarily because of the enormous percentage of lecturing time that TAs spend exhibiting "T" behavior. If TAs exhibited more complicated behavior in lecture, the protocol would have to be revised to reflect that.

3.1.3 IR:SQ Protocol

TA interactions with students during class execution and discussion sections are analyzed with a protocol called IR:SQ, which was designed by the author of this dissertation to track TA reactions to student stimuli. While the IR:SQ protocol is being used, the observer surreptitiously follows the TA around the room and listens to the overall tone of the his or her comments during interactions with the students. There are four possible categories into which TA comments can be placed:

1. **Initiate : Statement** is the category that represents the TA making declarative statements to students which are not, in any way, indicative of a response to a request for information or help from a student. TA behavior is marked as I:S if the TA approaches a group and informs his/her students of some fact or issues direct instruction. TA behavior can also be scored as I:S if the TA ignores a student question and proceeds to say something that is evidently unrelated to the student's request.

2. **Response : Statement** is representative of a TA answering a student's question directly, without forcing the student to reflect on the question at all or answering with a question of his/her own. No other behavior is coded as R:S behavior.

3. **Initiate : Question** is the category that represents the TA asking a student or

students a question that is not prompted by something a student has said. This behavior most often represents a TA approaching a group and asking a question but also represents a TA changing the topic of a conversation using a question.

4. **Response : Question** indicates that the TA has responded to a student question with a question. This could be as simple as asking a student to explain her or his question further, or as elaborate as asking the student a leading question designed to help the student answer his or her original question. If the TA responds to a student question with a statement designed to cause the student to work out the problem, then that is coded as R:Q even if the wording of the TA's sentences is grammatically declarative, not interrogative. Likewise if a statement is interrogative, but is intended to be, or ends up being, a rhetorical question, then that would be scored as R:S.

This protocol is not timed, and the observer does not focus on anything other than the characterization of what the students and the TA are saying. The TA receives a unique score for every behavior that either begins interaction with students, or else immediately follows a student speaking. The IR:SQ protocol is designed this way to measure how TAs react to or open conversations with students. For example, if TAs respond to student stimuli by trying to force students to arrive at their own conclusions, then their total scores should be weighted towards the Response : Question category. Likewise, if TAs respond to student questions by immediately providing the students with an answer their R:S category should be very high.

3.1.4 Time-Motion Study

As used in Spring 2004

In the revised time-motion study protocol, the TA's movement about the room was tracked, noting how many seconds he/she interacted with a given group, whether the TAs initiated the contact or responded to a student's request, and in what order the groups were visited. In the case of the graduate students assigned to teach the lab on E&M and circuit design, whether the TA manipulates the equipment was also noted.

A drawing of the room and the distribution of student groups was recorded, as well as the genders and names of each group member. When a TA stopped by a group an x was marked for every fifteen seconds that the TA was interacting with the group (rounding up for overcounting). When the TA left the group the cluster of x's connected to that visit was circled and a number written on the edge of the border to record the order in which TAs traveled around the room. This is a timed protocol so it is possible to evaluate the amount of time that a TA spends *not* interacting with any group.

Fall 2004

Beginning in the fall of 2004, it was noted whether a TA was responding to a student request or initiating contact with a student in situations when he or she entered into conversation with a group. This was done by marking an **I** or **R** next to each individual interaction with a group as soon as a TA initiated contact. This was done to determine if TAs interacted with groups for a substantially longer period of time depending on they initiated the discussion.

Spring 2005

In the Spring of 2005, the protocol was altered so that if the TA spent 15 seconds speaking without any of his or her students speaking, a T was marked next to the group. If the student spoke within these 15 seconds, then that time was scored as S instead of a T. In this way, a lower limit on the amount of time a TA spent speaking while conferring with a group could be established. This observation is valuable insofar as TAs should spend most of their time listening to students speak and occasionally prompting them to reflect on the physics involved in the discussion in new ways. Before the implementation of this revision of the time-motion study no mechanism was in place to assess the amount of time that TAs spent actively directing discussions.

3.2 Weekly Quizzes

In order to assess student learning during labs, quizzes were given the week after the lab was completed. These quizzes were designed to do one of three things:

1. Test a concept that students were exposed to or were intended to deduce during the course of the lab.

2. Analyze student retention of the basis for or execution of the lab.

3. Probe students' knowledge of concepts not explicitly covered in lab to determine if additional laboratory work is required to help establish theoretical understanding.

The quizzes generally included between three and seven questions, and were designed to be completed in the first ten minutes of class time. They were given no sooner than one week after the corresponding lab was completed.

3.3 TA Training

3.3.1 Summary of TA Training Research

While TA training was not a major part of this study, some changes were made to the methods of TA instruction in the Spring of 2005. This section summarizes some articles written on issues surrounding TA training and the methods of effectively implementing such training. A review article [78] was published in 2000 on issues pertaining to all aspects of graduate student education, including TA training, and may be of some assistance to the reader.

Stanford University [79] has undertaken a self-examination of their different departments with an eye towards finding the key features of departments that appear to be successful in producing high quality TAs. This study was qualitative and included each and every college at Stanford, not just the science departments, so these results may not be applicable to all science departments. The study's committee concluded that TA training appears to be successful when one or more of the following structures is in place to support TAs:

- Orentation/Training seminars at the beginning of the year each quarter for new TAs
- A pedagogy course or opportunities for ongoing discussion with peers and faculty during the first year of teaching
- Mentor TAs or peer mentoring systems
- A customized departmental TA handbook, online or in print
- Opportunities for practice or simulated teaching

- A midterm or “formative” TA evaluation, as well as an end-of-term or “summative” evaluation
- An archive TA training materials and courses
- Customized presentations by CTL (Center for Teaching and Learning) staff

Studies focusing directly on effective TA training for graduate students in physics [80] speak to the need for continuous graduate student assessment, videotaping graduate students teaching labs, review of videotapes with qualified mentors, and a variety of other techniques that can be used to support graduate students throughout their teaching tenure.

These articles were read prior to the rewrite of the second semester physics lab and the retraining of the TAs, however practical considerations of what the laboratory supervisor would allow determined the form of the retraining, thus techniques which have been shown to be effective through research were not used.

3.3.2 TA Training as Implemented in 2005

At present, TA training is confined to a series of five training sessions of two hours apiece administered just before and during a graduate student’s first year in the physics department. These sessions involve reading articles pertinent to laboratory education, lectures from professors in the physics department, and have recently begun to include some role-playing. Results of observations from previous semesters indicated that this training was not effective in encouraging TAs to administer their classrooms through Socratic dialogue.

Hake [81] published results from labs administered entirely through Socratic Dialogue, called SDI (Socratic Dialogue Inducing-Labs) and reported good success. Hake designed his labs in such a way that they aren’t entirely TA dependent by inserting instructions for the students to engage in Socratic debate or perform specified tasks. Thus the SDI laboratories are not as TA-dependent as the physics laboratories developed at the University of Arizona.

The format that was followed in TA training in the Fall of 2005 was that half of the time in lab meetings was given over to one of three tasks:

1. Review of discussion section problems and/or appropriate prompting questions to administer during the lab.
2. Role-playing discussion sections, with two of the three course TAs working through problems together and the third TA attempting to guide them through questioning.
3. TAs demonstrating lectures that they intended to give in their classrooms and critiquing them.

The author attempted to orchestrate the lab meetings for the rewritten laboratory so that roughly equal time was spent on all three activities over the course of the semester; however, it is difficult to determine whether this actually occurred, as this segment of the study was not monitored by an independent observer.

Qualitatively, activities 1 and 2 seemed to suffer greatly from the author's status as graduate student rather than instructor. The TAs were not used to accepting instruction from another graduate student, and did not appear to take either role-playing the discussion questions or the discussions occurring during the TA meetings very seriously.

TAs appeared to react positively to modelling their lectures for each other. This may have been due to the fact that the majority of the criticism and comments that TAs received for their lecturing style were given by other TAs, and thus it was more of a cooperative effort.

Training in the use of wait time, as described is section 2.3, was accomplished through the handing-out of one of Rowe's more accessible articles [70] and urging the TAs to actually count to "four" or "five" in their heads after posing a question and after receiving an answer.

3.4 Interviews

Two different interview protocols were used in this study. Both sets of interview protocols were conducted according to the guidelines laid out is section 2.6.3. The first interview protocol was used during spring of 2004 and is as follows:

1. What did the student feel were the lab's strengths, weaknesses, things liked, things disliked, things that should be changed?

2. In the initial whiteboarding sessions, was the student usually the group leader, very quiet, feel like he or she had things to contribute, and/or was interacting well?
3. What was the student doing during experiments? Data collection? Data transcription? Manipulating the equipment? Setting up the equipment?
4. What was the student's current lab skill level? Could the student articulate the techniques involved in data analysis? Measurement?
5. Why is the student in physics? Did the student have long-term career goals in physics? What were the student's opinions about physics?

The second protocol was implemented in the fall of 2004 and spring of 2005 in order to reflect changing research questions. The new protocol included the following questions:

1. What was the student's major and reason for being in the physics class?
2. Did the student have anything he/she would like to say before we begin? Anything he/she feels particularly strongly about?
3. Did the student use the consultation room during this semester?
4. Did the student feel that the discussion sections were of benefit? Why or why not?
5. Did the student feel that he/she constructed their own experiments during class?
6. In how large a group did the student normally work?
7. What was the student's role in the group? How often did the roles change?
8. What could the student remember from lab? (Prompt students to recall information from a lab if necessary.)
9. What did the student gain from doing the lab writeup that the student didn't know by the end of class?
10. What was the student's expected grade?
11. Did the student learn error analysis?

12. How would the student improve the labs?

13. Would the student like to add anything?

CHAPTER 4

RESULTS

The laboratory curricula studied in this dissertation will be referred to as the "revised laboratory curricula" and the second semester laboratory rewritten by the author of this dissertation will be referred to as the "rewritten laboratory." The analysis of laboratories will be divided into the phases a laboratory passes through during the course of a class: those sections being Discussion Section, Lecture, Class Execution, and Closing Discussion.

4.1 Study of Revised Laboratory

4.1.1 Overview of Section

In this section the discussion section of each lab, the lecture, the style that TAs use to administer their laboratory, and different approaches to closing discussions will all be presented and analyzed. For overviews of the observation techniques used please refer to section 3.1.

The discussion section was analyzed by tracking the number of questions posed to the students and the total amount of time that students spent in the different phases of the discussion (receiving question, working on question, presenting solution). Lectures were timed and analyzed with the SATIC in the fall of 2004 and with the ROP in the following semesters so as to create a picture of how TAs were using their time.

While the TAs were supervising the classroom by walking around and interacting with groups, their activities were monitored with time-motion studies in order to generate a picture of how they spent their time. This information was vital in determining how TAs engaged with students, and was coupled to data collected using the IR:SQ protocol in order to determine if TAs administered their classes through Socratic dialogue.

4.1.2 Description and Naming of TAs

In order to protect the anonymity of the different TAs who have been involved in this study Greek letters were used to replace their real names throughout this document. In the event that a TA was in the study for more than one semester, he or she was given a different descriptor for each independent semester.

4.1.2.1 Spring 2004

- **Alpha** (α) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately four years and taught one section of the second semester laboratory during the spring of 2004.

- **Beta** (β) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately two years and taught one section of the second semester laboratory during the spring of 2004.

- **Gamma** (γ) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately three years and taught one section of the second semester laboratory during the spring of 2004.

- **Delta** (δ) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately three years and taught one section of the fourth semester laboratory during the spring of 2004.

- **Epsilon** (ϵ) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately three years and taught two sections of the fourth semester laboratory during the spring of 2004.

4.1.2.2 From Fall 2004

- **Zeta** (ζ) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately four years and taught one section of the third semester laboratory during the fall of 2004.

- **Eta** (η) is a foreign female who has been a graduate student at the University of Arizona for approximately three years and taught two sections of the third semester laboratory during the fall of 2004.

- **Theta** (θ) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately two years and taught two sections of the first semester laboratory during the fall of 2004.

- **Iota** (ι) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately one year and taught one section of the first semester laboratory during the fall of 2004..
- **Kappa** (κ) is a female U.S. citizen who has been a graduate student at the University of Arizona for approximately five years and taught two sections of the first semester laboratory during the fall of 2004.

4.1.3 Study of Discussion Section

The discussion section begins immediately after the start of class and consists of a period of time during which students work out problems in small groups on whiteboards. Different TAs treated discussion sections very differently, thus the data presented in this section is qualitative, and occasionally reports that TAs do not hold discussion sections at all.

In classes where discussion sections were held, they were studied and timed, with the timing being broken down into three different parts:

- The posing of the discussion question, when a TA or student is writing the question on the board.
- The group discussion and whiteboarding session, when students are working in groups on the problem.
- The presentation part of the discussion session, either the TA or the student presents the solution to the problem.

4.1.3.1 Spring of 2004

Alpha, Beta, and Gamma

No data exists on Alpha's (α) handling of his discussion section. Due to a class conflict he wasn't able to be present for the beginning of the discussion section and consequently the author handled that part of the class. Beta (β) only had a formal discussion session during the first lab. By the second lab he had forgotten to have the discussion session at the start of class and instead tried to shift it to the end. He repeated this on the third lab and then abandoned the discussion section entirely.

Gamma (γ) held discussion sections intermittently, and his discussion sections were marked by his working out problems for students on the blackboard when he noted that students were encountering difficulties with the problems assigned. On those occasions when students would ask questions related to homework Gamma would either solve the problem immediately on the board or else attempt to engage in a dialogue with the class as a whole in which he would question them about how to approach the problem. When Gamma administered a discussion section in this manner the SATIC (as measured by two observations) looked like:

- Initiatory (Talking): 55 %
- Initiatory (Questioning): 16 %
- Responding (Teacher-Centered): 25 %
- Responding (Student-Centered): 4 %

This data indicates that even when Gamma was trying to engage his students in dialogue about a problem he still spent over 50% of his time speaking directly to them, and student-centered questioning and responding only took place 20% of the time. This indicates that these sections are of dubious benefit to students as they are being asked to demonstrate understanding or comment on the questions a small percentage of the time.

Delta and Epsilon

Neither Delta (δ) nor Epsilon (ϵ) used discussion sections on a regular basis to administer the third semester labs, preferring to step immediately into lecture and then begin the lab.

4.1.3.2 Fall of 2004

In fall of 2004 the discussion sections were analyzed much more carefully in order to learn how TAs were spending their time. Discussion sections were required by the course instructors, so ample data exists for analysis.

The observations of the discussion section were divided into the three segments previously described. The results from discussion sections are presented in Table 4.1. The observations taken in this semester show clearly that TAs held discussion sections that lasted approximately half an hour, and that most TAs had difficulties in addressing several questions over that period of time. Theta (θ), Iota (ι), and Kappa (κ) all held discussion sections on a regular basis and elicited the discussion questions from the students in the lab if possible.

Category	Zeta (ζ)	Eta (η)	Theta (θ)	Iota (ι)	Kappa (κ)
Pose Discussion Question	2 m 25 s	25 s	1 m 38 s	2 m 23 s	1 m 30 s
Group Discussion	19 m 10 s	18 m 1 s	9 m	3 m 15 s	8 m 51 s
Presentation Length	2 m 17 s	3 m 3 s	8 m 40 s	2 m 30 s	3 m 6 s
Number of Questions	1.1	1.4	1.4	3.7	1.4
Total Length	26 m 15 s	29 m 19 s	27 m 1 s	30 m 5 s	18 m 50 s
No. Sections Analyzed	5	4	9	6	7

Table 4.1: Analysis of Discussion Section in Fall 2004

Theta

Theta's (θ) style of managing a discussion section was to focus on making sure that students worked carefully through the problems and that all the students in the classroom understood the work being done, which led to him working through fewer problems per lab than perhaps he could have. His class was notable for the long presentation time of the students, which was brought about by Theta (θ) asking detailed questions of the students doing the presentation. These questions appeared to provoke serious reflection on the part of his student and could have been of considerable benefit.

Iota

Iota's (ι) students worked through an average of 3.7 problems per discussion section, far more problems than students in other labs. Iota's time-motion studies indicated a relatively low incidence of interaction with students while students were doing work. This may have helped him keep his students working out problems with great rapidity. The time that his students were allowed to complete problems was generally adequate for them, however the fact the Iota (ι) was not questioning his students while they worked on whiteboards means that there were few controls in place to make sure that students adequately explored the problems before they were called to present.

Kappa

Kappa drew the discussion questions used in her class exclusively from student suggestions, so if students lacked suggestions then discussion sections were not held. Generally only one problem was attempted and she focused on helping students complete in in a reasonably expeditious fashion. This explains the relatively short length of her discussion section, which, at 18 minutes and 50 seconds, is by far the shortest of all the sections considered in this study.

Zeta

Zeta's (ζ) style of instruction was to assign students long discussion questions which, in a little less than half of the instances observed, he would end up working on the board himself after students had failed to finish them. Zeta allowed his students a large amount of time (an average of 19 minutes) to complete the questions, however he lacked training to guide his students to complete the problem with his help.

Eta

Eta's (η) teaching style involved beginning a discussion section with a question of her design, and then adding additional questions as some groups of students finished the first question. This approach appeared to engender confusion in the discussion section, and made it difficult to sort out the amount of time that students spend working on any one particular problem. Eta's (η) students worked through anywhere between one and four problems per discussion session due to the assignment of multiple problems. Many of these problems were worked in parallel and it's not clear that any groups were able to complete all of the work assigned. A typical discussion section could include an initial question and up to three follow up questions asked during the course of the discussion section. In general Eta (η) would work out the majority of the additional questions herself and only hold students accountable for presenting responses to the initial question.

4.1.3.3 Conclusions from Study of Discussion Section

Observations of the TAs directing discussion sections revealed that TAs were often unprepared for discussion section insofar as they did not bring appropriate questions to class, instead relying almost exclusively on student suggestions of what problems to work. This resulted in some discussion sections being extremely short (with 19 minutes being a lower end) and other discussion sections lasting for closer to half an hour but only addressing one problem. Both states are not reflective of a long discussion section in which many different problems are considered.

After studying how TAs conducted discussion sections, it was decided that the most effective way to help TAs conduct discussion sections in which a reasonable number of problems were covered would be to provide TAs with problems of an appropriate level of difficulty. Also provided were the solutions to these problems and instructions to the TAs that they should interact with groups through Socratic dialogue while the discussion section was going forward.

It was anticipated that this would lead to TAs having active discussions with students while they were working on discussion problems, potentially exposing student misconceptions and aiding in establishing an attitude of serious and focused work throughout the discussion section.

4.1.4 Analysis of Lectures

Physics laboratories taught at the university often include the TA giving a short lecture concerning the topics to be taught in the lab. Traditionally the TAs are not given a list of appropriate topics to be covered and frequently are not aware of the students' progress in the associated lecture course, so their lectures are improvised from their own sense of what is appropriate and inappropriate to cover in lab.

In addition the TAs seldom introduced the labs in the context of the premise presented in the lab manual. During interviews different students made comments indicating that they saw little connection between the lab and the premise of the labs, such as the interview in the following excerpt from an interview from the fall of 2004. The interviewer is identified with the letter I, and the student with the letter E.

- I: Would you say that the labs are clearly written and you're able to decipher the purpose of them?

- E: The lab book is kinda...we learned very quickly not to rely on it what-so-ever. We, uh, the lab book, uh, usually the lab instructor just decides to completely use a different laboratory procedure than what is in the lab book so it's pointless anyways.

A partial explanation as to why the lab TAs did not incorporate the theme of the lab in their lectures may be seen in Kappa's (κ) reporting that she does not read the TA notes because she has taught the labs so many times that reading the manual is unnecessary. She has never taught a section of the revised labs before, and hence it is unsurprising that she conducted her laboratory for physicists similarly to how she administers her engineering laboratories. The relative success of her students on the quizzes speaks well to her ability to teach, or to the ability of her students, but she never used the revised teaching techniques that the TA notes called for.

4.1.4.1 Spring of 2004

In the spring semester of 2004 the TAs' lectures were studied using the SATIC protocol as described in section 3.1.1. The SATIC results for those TAs are presented in Table 4.2. Note Init indicates Initiating and Resp indicates Responding.

Category	Alpha(α)	Beta (β)	Gamma (γ)	Delta (δ)	Epsilon (ϵ)
Lecture Length	6 m 48 s	46 m	41 m	12 m 30 s	13 m
Init: (Talking)	63 %	43 %	85 %	98 %	100 %
Init: (Questioning)	25 %	26 %	10 %	0 %	0 %
Resp: (Teacher Centered)	13 %	29 %	5 %	2 %	0 %
Resp: (Student Centered)	0 %	2 %	1 %	0 %	0 %
No. of Lectures Tracked	7	5	6	5	4

Table 4.2: SATIC Analysis of Lectures Given in Spring of 2004.

Alpha

Alpha (α) had a very informal lecture style. Over the course of the entire semester his lectures lasted for an average of 6.8 minutes at the beginning of lab, and this time was used to frame the context of the lab and to present a series of problems for the students to think about. He would then release the students to work in groups on whiteboards, designing the experiment on their own. The SATIC results for Alpha (α) paint a picture of a TA who not only gives a short lecture but also makes his lecture very interactive, questioning students 25% of the time.

Beta

Beta (β) is the only TA who employed two different teaching formats throughout the semester, as he switched his teaching style mid-semester after discussions with the professor in charge of his class and the classroom observer.

During the first part of the semester Beta's (β) teaching style focused on having extremely long lectures (46 minutes) at the beginning of class. During these lectures Beta (β) would speak about the theory behind the lab and would spend 26% of his time questioning students. Many of his questions ended up being rhetorical in nature as he was reluctant to force students to answer his questions and he didn't grant his students adequate wait time (section 2.5.2) to ponder some of them (If a question ended up being rhetorical in nature then that behavior was coded as talking, rather than questioning). After the lectures were concluded his students were allowed to begin the lab, although Beta would stop the class and engage in more lecturing if he felt that students were confused.

After conversations with the course instructor and classroom observer, Beta (β) changed his lecturing style for the labs thereafter. He would lead in with a four minute lecture and then release the students to "play" with the equipment. He did, however, still stop the class when he felt it was necessary to have a group discussion, which were still very teacher centered. A SATIC analysis of such a thirty minute long discussion follows

- Initiatory (Talking) 68 %
- Initiatory (Questioning) 10 %
- Responding (Teacher-Centered) 18 %
- Responding (Student Centered) 3 %

These mid-class discussions were even more teacher centered then the opening lectures because they were convened to fix problems that Beta perceived were occurring with the lab. As such he tried to get all the students to address these problems and, again, often wasn't very successful in doing so. The discussions would drag into long silences as the students digested Beta's (β) comments and attempted to formulate responses. Beta (β), growing impatient, would answer his own questions and move on to ask more questions which the students didn't readily comprehend. Situations in which Beta (β) would answer his own questions were marked as Initiatory (Talking) behavior to reflect the fact that, even though the students may have been reflecting, there was no actual interaction. This is consistent with the SATIC format which includes rhetorical questions as Initiatory (Talking) behavior.

Gamma

Gamma (γ) tended to organize his lectures around fairly complicated derivations of equations important to the lab, thus his lecturing style tended toward direct transfer of information (85% of his time was spend talking) and seldom involved questioning the students (10% of time spent questioning). As the derivations of equations Gamma (γ) considered important to the laboratories could take a considerable amount of time, his lecture length is much longer than the average observed throughout this study, averaging 41 minutes.

Delta

Delta's (δ) lectures were focused on delivering information to students in the shortest amount of time possible. His lecture length was 12 minutes and 30 seconds, and 98% of his time was spend directly imparting information. The technique used to observe the TAs' lectures is insensitive to variables such as number of words a minute and/or amount of information contained in a lecture, however qualitative observations indicated that both of these variables were extremely large. Delta (δ) began a lecture by focusing on the underlying theory of the laboratory and then concluded by writing a list of topics to be investigated on the board.

Epsilon

Epsilon (ϵ) taught his lecture in a similar fashion to Delta (δ). The average lecture time was thirteen minutes and 100% of his time was spend directly imparting information. Any questions that were asked Epsilon (ϵ) invariably came after the lecture was complete or during the experimental part of the lab. Epsilon (ϵ) would demonstrate equipment for students during his lecture, interlacing demonstration of the equipment with the lecture rather than having a separate experimental section.

4.1.4.2 Fall of 2004

Beginning in the fall of 2004 the SATIC protocol described in section 3.1.1 was no longer used for observations of lectures and the ROP described in section 3.1.2 was used instead. The results for TAs studied in the fall of 2004 are presented in table 4.3

Recall that the ROP is designed so that it overcounts student-centered teaching behavior, thus the ROP results for Eta (η) could be read as, "Eta (η) spends at least 34%

Category	Zeta (ζ)	Eta (η)	Theta (θ)	Iota (ι)	Kappa (κ)
Lecture	13 m 18 s	21 m 32 s	22 m 55 s	15 m 45 s	4 m 8 s
Talking	43 %	34 %	40 %	49 %	41 %
Whiteboarding	39 %	28 %	17 %	8 %	13 %
Demonstrating	15 %	7 %	1 %	9 %	16 %
Teacher Centered	**97 %**	**69 %**	**58 %**	**66 %**	**70 %**
Questioning	1 %	19 %	36 %	24 %	19 %
Answering	2 %	10 %	4 %	4 %	10 %
Students whiteboard	0 %	2 %	2 %	6 %	0 %
Student Centered	**3 %**	**31 %**	**42 %**	**34 %**	**29 %**
No of Lectures Tracked	5	7	6	4	4

Table 4.3: ROP Analysis of Lecture for TAs Studied in Fall 2004

of her time lecturing without writing on the whiteboard or demonstrating equipment, at most 28% lecturing while writing on the whiteboard, and at most 7% of her time lecturing while demonstrating equipment. This means that the percentage of her lecture that is teacher-centered is at least **69%**. Eta (η) spends her lecture time questioning students at most 19% of the time, answering students' questions at most 10% of the time, and having students work out problems on the whiteboard at most 2% of the time. This brings her student centered time to at most **31%**.

These observations reveal dramatic differences in the lecturing styles of the TAs assigned to the revised laboratories. Kappa (κ) lectured for only an average of 4 minutes and 8 seconds, whereas Theta (θ) lectured for 22 minutes 55 seconds, indicating a lack of uniformity in lecture length. Lecture content varied smoewhat as well, with one TA, Zeta (ζ), focusing on lecture as a method of delivering information without student response while other TAs interacted with their students during lecture. Some of the TAs studied in fall of 2004 questioned students during lecture an extensive amount, as can be seen from the results presented in Table 4.3. Of the TAs studied that semester, Theta (θ) and Iota (ι) both tried to have the students establish the context for the experiment by questioning the class as a whole, as evidenced by the fact that Theta (θ) was questioning for over 36% of his lecture and Iota was questioning for over 24% of his lecture. The fact that some TAs were using their lecture to question students suggests that a study should be conducted to assess the effectiveness of this behavior.

The percentage of the class that participated in these class-wide discussions during lecture was recorded and averaged over the semester. Also recorded was the percentage of the discussion that was contributed by the most vocal and second most vocal speaker in the class. The method used to identify these students was to count the number of responses to TA questions and/or the number of times these students initiated discussion on their own and consider these two numbers added together an accurate measurement of classroom participation.

Theta (θ) was able to involve, on average, 45% of his class in answering at least one question during lecture. Of those 45%, the most vocal student in any given class was responsible for, on average, 33% of all comments made and the second most vocal student in any given class was responsible for, on average, 17% of all comments made. Thus just two students made 50% of all student comments. These data were drawn from 134 questions tracked during nine different sampling periods.

Iota (ι) was able to involve, on average, 44% of his class in answering at least one question during lecture. Of those 44%, the most vocal student in any given class was responsible for, on average, 35% of all comments made and the second most vocal student in any given class was responsible for, on average, 25% of all comments made. Thus just two students making 60% of all student comments. These data were drawn from 94 questions tracked during five different sampling periods.

This paints a picture of a classroom in which the TA is often asking questions, but the questions are answered by only a few very vocal students. While this is undoubtedly beneficial to the students who are engaged in discussion, it is doubtful if this is of benefit to the rest of the students in the classroom.

4.1.4.3 Conclusions from Analysis of Lecture

The data indicating that TAs who have long lectures are unsuccessful in engaging the entire class using questioning is suggestive that questioning students during lecture is not a successful method of instruction in this context. Furthermore, qualitative observation of students revealed signs of disinterest after fifteen minutes, with some extreme examples including students laying their heads down and sleeping during lectures.

It was thus decided that when the second semester laboratory was rewritten, considerable emphasis would be put on having TAs give lectures that are as short as possible, if they chose to lecture at all.

4.1.5 Lab Execution

4.1.5.1 TA Interaction Time

TA behavior was analyzed through a protocol called the modified time-motion study (described in section 3.1.4). This protocol measures how much time that TAs spend working with students. It also indicates whether TAs initiated the contact without being asked or responded to students' requests for help.

Spring 2004

The first semester this study was conducted there were relatively few time-motion studies conduced because the observer was focusing closely on the content of TA interactions rather than generating an aggregate statistical picture of TA activities. The time-motion study was not fully fleshed out at this point, and hence information such as if TA was initiating contact or responding to students was not recorded in the spring of 2004.

Alpha's (α) average interaction time with students after the beginning of the lab (during walking around asking questions or responding to questions) is on the order of one minute, seven seconds per group, indicating that on average he did not pause long with any one particular group. This conclusion is based on three time-motion studies over the course of the semester. Beta (β) spent an average of one minute, five seconds with groups when he interacted with them while the actual lab was in progress. This conclusion is based on four time-motion studies over the course of the semester.

Gamma (γ) interacted little with the students outside the medium of lecturing in front of the class. Students would occasionally approach him and ask questions, which he answered; however he did not initiate contact with groups as a general rule. After spring of 2004 the time-motion protocol was revised so that when students approached TAs and spoke to them it was counted as an interaction. Delta's (δ) average interaction time was 49 seconds. This conclusion was reached from three time motion studies.

Epsilon's (ϵ) style of interacting with the students was to wait for them to ask him questions rather than questioning the students, thus his interaction profile was not measured. After spring of 2004 the time-motion protocol was revised so that all times when students approached TAs and spoke to them it was counted as an interaction.

Fall 2004

In the fall of 2004 the TA behavior was analyzed much more carefully. Table 4.4 represents all of the time motion studies that were conducted during that semester. The results for Zeta (ζ), for example, can be read thusly from the table: Zeta (ζ) spends at most 76 % of his time speaking to his students. He initiates contact with his students 31 % of the time, and responds to requests for help 69 % of the time. When he initiates contact he takes, on average, one minute, eight seconds with each group. When he responds to request for help he takes, on average, 58 seconds with each group. This data is based on of seven different time-motion studies conducted over the course of the semester.

Category	Zeta (ζ)	Eta (η)	Theta (θ)	Iota (ι)	Kappa (κ)
Max Engagement	76 %	63 %	88 %	38 %	29 %
Initiate/Respond	31%/69%	70%/30%	69%/31%	47%/53%	44%/56%
Initiate Time	1 m 8 s	30 s	45 s	34 s	36 s
Respond Time	58 s	41 s	1 m 21 s	1 m 2 s	38 s
No. Time-Motion	7	11	13	11	12

Table 4.4: Time-Motion Studies for TAs in the Fall of 2004

Conclusions from Time-Motion Studies

The observations conducted in the fall of 2004 provide a clear picture of how TAs spent time in their lab. Iota (ι) and Kappa (κ) allowed much of their time to go unused in lab as seen by their low total amount of time engaged (38% and 28%, respectively). A laissez faire style of administering a classroom is antithetical to the principles of a Socratically taught lab, for which a TA should have a series of probing questions predesigned and should use these questions to confirm that students are developing understanding of the material.

Zeta's (ζ) behavior is marked by a high engagement time (76%), however close analysis of his behavior reveals that he is responding to requests for assistance much of the time instead of initiating contact to monitor student progress (responding 69% of time). Other TAs, such as Theta (θ) and Eta (η) have relatively high engagement (88% and 63%, respectively) and also frequently initiate contact (initiating 69% and 70%, respectively), indicating that they may be properly monitoring their classrooms. This data overall paints a picture in which several TAs could not possibly be adequately monitoring their classrooms due to their low engagement time, which requires correction.

When the second semester laboratory was rewritten special attention was paid to providing TAs with effective questions that could be used to probe student understanding during the course of the lab and the discussion section. It was emphasized to all TAs that they should interact continually with their students during the lab.

4.1.5.2 Contact with Equipment

Fall 2004

The TAs in charge of the third semester laboratory were specifically analyzed to see how often they made deliberate and significant contact with the students' equipment. This included rewiring a circuit, changing the settings on an oscilloscope so that a signal was visible, or doing anything else that dramatically aided a student. The reason that this analysis was conducted is that the third semester labs tended to only have a few experimental setups per experiment. If the TA assembled a circuit for the students, then he/she has did a substantial portion of the students' work for them.

1. Zeta (ζ) manipulated the students' equipment roughly 52 % of the time that he interacted with a group. This was established by tracking his actions in six time-motion studies.

2. Eta (η) manipulated the students' equipment roughly 45 % of the time that she interacted with a group. This was established by tracking her actions in five time-motion studies

Thus both TAs had a problem in that they changed the wiring, oscilloscope settings, or otherwise manipulated equipment when students should be required to complete their own equipment setup.

Conclusions from Study of Contact with Equipment

Verbal warnings were included with the rewritten lab manual asking TAs not to manipulate students' equipment.

4.1.5.3 TA Response Profile

Both modern advocates of reform in physics education and those who favor more traditional classroom instruction have as part of their educational culture the belief that students do not benefit in any way if they are simply given an answer. Teaching through questioning has been taken as a sign of high-quality instruction since the rise of Greek and Chinese philosophy, and has remained an educational mainstay throughout many reforms. Thus TAs should not simply give answers to questions, and will instead make students work for their own edification. Some TAs are much better at forcing students to work and think in class, and the challenge is to eliminate the behavior of simply giving students answers to questions.

The IR:SQ profile described in section 2.1.3 was used to track TA behavior during interaction with groups. Here the questioning behavior (both initiate: question and respond: question) is indicative of a TA who administers group work through Socratic dialogue.

Fall 2004

Category	Zeta (ζ)	Eta (η)	Theta (θ)	Iota (ι)	Kappa (κ)
Initiate : Statement	35 %	28 %	26 %	26 %	21 %
Response : Statement	51 %	37 %	18 %	51 %	48 %
Total Declarative	**86 %**	**65 %**	**44 %**	**77 %**	**69 %**
Initiate : Question	9 %	26 %	40 %	13 %	10 %
Response : Question	4 %	9 %	16 %	9 %	20 %
Total Inquisitive	**13 %**	**35 %**	**56 %**	**22 %**	**30 %**

Table 4.5: IR:SQ responses for TA's in Fall of 2004

The results for the fall semester of 2004 are collected in Table 4.5. Note that Kappa (κ) responded to students inquiries with a question 20% of the time, and Zeta responded with a question only 4% of the time. Kappa (κ), a highly seasoned TA, was much better at not "giving away" information to students than Zeta (ζ), but they both received TA training at about the same time.

In general all of the TAs studied in the fall of 2004 were not good at engaging students in Socratic dialogue, with the exception of Theta, whose interactions with students fell into the inquisitive categories over 50% of the time. Since Socratic dialogue requires both extensive interaction with students and use of questioning to lead students to form their own conclusions, only Theta (θ) could have had a classroom governed by Socratic dialogue, since the only other TAs to interact extensively with their students, Eta (η) and Zeta (ζ) did not show interactions governed by questioning.

Conclusions from TA Response Profiles as measured by IR:SQ

Some TA's, such as Zeta (ζ), showed a remarkably low incidence of asking their students questions and/or soliciting results. This indicates that Socratic dialogue is not being used in these labs. It was decided that in the rewritten laboratory TAs would spend time in their weekly meetings practicing reacting to students by using questioning as a method of leading students to correct conclusions, as opposed to simply providing information.

4.1.6 Closing Discussions

Every TA had a different method of closing lab. Their styles of closing class are briefly summarized in this section.

4.1.6.1 Spring of 2004

In the spring of 2004 none of the TAs studied had their students do closing discussions. Alpha (α) and Beta (β) would be in periodic contact with groups spread throughout the room and would allow them to leave after they had completed the labs to the TAs' satisfaction. Gamma (γ) had a slightly different approach, in that students had to have him sign their lab manuals before they left, and this was the time he questioned them to make sure that they had an adequate amount of data or that they understood said data. Unfortunately, no timing was done of this activity, but informal estimation indicated that Gamma signed students out in a matter of seconds while glancing at their lab manual or questioned them for upwards of forty seconds if he felt that their lab manual looked suspicious. The classroom observer never saw Gamma send students back to gather more data.

Delta (δ) and Epsilon (ϵ) both looked over their students' data and signed them out when they completed the laboratory.

4.1.6.2 Fall of 2004

The TAs teaching in the fall of 2004 all also had their own style of dismissing class. Both Zeta (ζ) and Eta (η) allowed their students to leave the lab after they self-reported that they were satisfied with the work they had done. Theta (θ) either had a closing discussion or simply dismissed students after they had finished their work. When he used a closing discussion, it was a directed whole classroom discussion that he moderated. This could be quite lengthy, with an average time of 33 minutes, much of which was comprised of student presentations. This number was established through observations of three of Theta's closing discussions. Iota (ι) always used a closing-class discussion, but tended to keep it short, with an average time of nine minutes, almost all of which was student presentations. This time was established through observation of four of Iota's closing discussions. Kappa (κ) simply allowed students to leave the lab after they reported that they were satisfied with the work they had done.

Note that many TAs did not capitalize on the opportunity to push their students to do and present their analysis of data in class. This is significant because inquiry based education is a difficult thing to oversee after the student has already left the room. In free-form labs with no set objectives, students can fail to make correct conclusions, or any conclusions at all, without supervision and feedback.

4.1.6.3 Conclusions from Study of Closing Discussion

It was decided that TAs would benefit from being provided with a list of potential closing discussion topics that they could share with their students in a manner of their choosing. TAs were instructed to provide students with a topic or to solicit students to develop a topic not later than within an hour of the closing of the lab so that students would have time to prepare a presentation and demonstrate understanding of the experiment.

4.2 Quiz Results

The weekly quizzes were used almost exclusively to inform how the labs should be rewritten. The questions were designed to probe the effectiveness of the individual laboratories, but have some validity as a tool to study the effectiveness of the laboratory curriculum as a whole. The quiz questions given each semester were sorted into conceptual and non-conceptual categories. Questions which were non-conceptual but did not test student understanding, such as, “Did you investigate this phenomena?” or “Do you recall the premise of the laboratory you did last week?” are not included in this analysis. If, for some reason, the quiz questions were not given to all of classes being studied then that question is exempted from this analysis.

The aggregate conceptual and non-conceptual quiz scores are reported in the following sections. The data were analyzed by computing the average quiz score and the standard deviation of the mean for each subsection (TA class and conceptual or non-conceptual items).

It is very important to remember that the number of students involved in this study is small, so the results presented here are not statistically significant and thus are proof of nothing. Ideally, different aggregate quiz scores on conceptual items between different sections would be indicative of the effects of different TA teaching styles, however with class sizes this small the distribution of students can have a significant effect.

4.2.1 Analysis of First Semester Laboratories

These laboratories focus on mechanics, and were taught by Theta (θ), Iota (ι), and Kappa (κ). Thirteen conceptual questions and thirty-five non-conceptual questions were given over the course of the semester, the results are displayed in Table 4.6.

	Theta (θ) sec 1	Theta (θ) sec 2	Iota (ι)	Kappa (κ)
Conceptual	53 ± 8	38 ± 9	33 ± 9	42 ± 6
Non-Conceptual	54 ± 5	59 ± 4	49 ± 4	46 ± 4

Table 4.6: Aggregate Quiz Scores ± Standard Deviation for First Semester Laboratories

Note that Theta's (θ) students conceptual score actually drops in his second section, indicating that in this instance they did not benefit from the TAs increased familiarity with the laboratory due to having taught the same lab immediately prior. This could be due to individual differences between students in the classrooms. Another interesting feature of this data set is the extremely low score of Iota's (ι) students, which could be due to his low instance of interaction with them leading to a lack of understanding of the underlying concepts of the lab.

4.2.2 Analysis of Second Semester Laboratories

These laboratories, as taught in spring 2004, included thermodynamics, fluid mechanics, waves and vibrations, and relativity. There were 19 conceptual questions and 36 non-conceptual questions given over the course of the semester, the results are presented in Table 4.7.

	Alpha (α)	Beta (β)	Gamma (γ)
Conceptual	72 ± 5	65 ± 7	70 ± 6
Non-Conceptual	56 ± 4	65 ± 5	61 ± 6

Table 4.7: Aggregate Quiz Scores ± Standard Deviation for Second Semester Laboratories

Here the students' scores in different sections are so similar to each other that it is impossible to speculate about potential differences between the classrooms by reasoning from quiz grades. All that can be said about individual classrooms is that Alpha's (α) students high conceptual score in comparison to their score on non conceptual items may reflect his concentration on student-oriented learning and exploration and lack of emphasis on having a long formal lecture.

4.2.3 Analysis of Third Semester Laboratories

The third semester labs focused on electricity and magnetism in general and circuit design in particular. They were assessed with 50 conceptual and 16 non-conceptual questions, with results are presented in Table 4.8.

	Zeta (ζ)	Eta (η) section 1	Eta (η) section 2
Conceptual	43 ± 4	48 ± 4	58 ± 4
Non-Conceptual	51 ± 8	32 ± 7	49 ± 9

Table 4.8: Aggregate Quiz Scores ± Standard Deviation for Third Semester Laboratories

Here Eta's (η) second lab sections quiz score is considerably higher than the quiz score associated with her first lab section. This could reflect individual differences between the classes or improvement in her teaching styles brought about by practice in the first section.

4.2.4 Analysis of Fourth Semester Laboratories

These labs centered around optics and modern physics. They were assessed using 26 conceptual questions and 16 non-conceptual questions. Epsilon's (ϵ) section 1 was not considered here as it contained only four participating students and was never attended by the observer. Quiz results are reported in Table 4.9.

	Delta (δ)	Epsilon (ϵ) section 2
Conceptual	57 ± 5	52 ± 6
Non-Conceptual	42 ± 8	27 ± 5

Table 4.9: Aggregate Quiz Scores ± Standard Deviation for Fourth Semester Laboratories

Both Delta's (δ) and Epsilon's (ϵ) style of classroom management and students' grades on quizzes are so similar that no useful conclusions can be drawn from this data. Epsilon's students low score on non-conceptual items is interesting, but not traceable to any behavior that protocols used are sensitive too.

4.2.5 Conclusions from Analysis of Quizzes

The quizzes were never intended to be used as a method of comparing the effectiveness of the different techniques employed by TAs; rather the questions were written to probe student misconceptions after instruction. It is *very* important to remember that the number of students involved in this study is insufficient to generate statistically significant differences. However the relatively low scores across the board reveal an overall lack of student understanding and indicates that there is a potential positive benefit to be had from lab revision.

4.3 Description of the Rewritten Lab

In this section of the dissertation the second semester laboratories written by the author and referred to as "the rewritten laboratory" will be analyzed using the same techniques that were used to examine the revised laboratory curricula. The discussion section of each lab, the lecture, the style that TAs used to administer the laboratory, and different approaches to closing discussions will all be outlined and explained. For overviews of the observation techniques used please refer to section 3.1.

4.3.1 Premise of Labs

There are a number of salient differences between the revised laboratories as originally envisioned, and the rewritten laboratories that were designed to correct weaknesses on the part of instructors and confusion on the part of students. Foremost among the changes implemented was the elimination of the short stories intended to establish context for the laboratories. Specific objections to and notes on each of these premises are provided in the suggested revisions document associated with this dissertation.

In the rewritten laboratory manual, students were told what experiments were available in the room and received some information about the experiments, but the techniques and mathematics necessary to do the experiment were not explained. In this way, students did not become frustrated by any difficulty in figuring out what the laboratory manual wanted them to do, and could focus their attention on designing experiments to accomplish the goals. This technique of directing students does, however, lack the advantage of introducing students to doing science by placing the labs in investigative contexts.

In order to correct for this deficiency, a short writing assignment, called a prècis, was written for each of the laboratories. These writing assignments were designed to be finished with an hour or two of work, and centered around the historical development of a particular branch of science, or the lives of particularly interesting scientists, or deeper reflection on scientific concepts.

4.3.2 Discussion Section

The TA notes contained sample discussion section problems with solutions to the problems for each lab. These problems were chosen to help introduce and stimulate discussion on issues important to the laboratory. After the middle of spring 2005 copies of the questions were distributed to the laboratory students so that TAs did not have to spend time writing the questions on the chalkboards. Several different sample questions were provided for each lab with the hopes that the discussion session would encompass four to five problems each week.

4.3.3 Lecture

Sample lecture notes were added to each copy of the TA notes in order to help the TAs design an appropriate lecture, and also to help TAs work through any misconceptions they may have concerning the physics of the lab. These lectures were not required; TAs were free to design their own or to skip lecturing entirely. The lectures were designed to take between ten to fifteen minutes each; however, timed trials put them in the range of fifteen to twenty minutes.

4.3.4 Execution

In order to assist the TAs in administering a classroom that is as Socratically based as possible, the TA notes contained sample prompting questions that could be used to assist students in both formulating a good experiment and working through technical difficulties in conducting their experiment.

4.3.5 Closing Discussion

The closing discussions were facilitated by a list of suggested topics included in the TA manual. TAs were instructed, both in the TA manual and in lab meetings, to give each group a different discussion topic at least an hour before the lab ended so that the groups would have time to generate appropriate graphs on their whiteboard and discuss the results they were going to present to the class.

4.3.6 Grading

Student interviews revealed substantial frustrations due to the perception of dramatically different grading styles among different TAs in the fall of 2004. The grading in the rewritten lab manual was standardized in the spring of 2005 using a grading rubric, which required that the labs be graded on a number of different categories. As long as TAs graded carefully using the rubric, the grading scale would be constant between different lab sections.

4.4 Study of Rewritten Laboratories

4.4.1 Description and Naming of TAs

4.4.1.1 Spring 2005

- **Nu** (ν) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately three years and taught one section of the second semester laboratory during the Spring of 2005.

- **Xi** (ξ) is a male U.S. citizen who has been a graduate student at the University of Arizona for approximately two years and taught two sections of the second semester laboratory during the Spring of 2005.

4.4.2 Discussion Section

Nu and Xi

Nu (ν) and Xi (ξ) were carefully studied throughout the course of the semester. Both TAs held discussion sections on a regular basis, had students whiteboard and finish problems, and required students to present the results of their work to the class. Observations in previous lab sections suggested that when one very hard question was the focus of the discussion section some groups of students would finish work early, whereas other groups would sometimes never finish their work. By assigning shorter and simpler problems it was hoped that the discussion sections would be of greater benefit to all. Thus Nu (ν) and Xi (ξ) were instructed to keep whiteboarding work in the vicinity of five minutes so as to allow students to complete a number of different problems, all of which were supplied in the lab manual and pertinent to the laboratory at hand. The behavior of TAs is presented in Table 4.10

Category	Nu (ν)	Xi (ξ)
Pose Discussion Question	39 s	1 m 18 s
Group Discussion	7 m 54 s	5 m 54 s
Presentation Length	4 m 40 s	3 m 22 s
Number of Questions	2.3	3.3
Total Length	30 m 23 s	34 m 52 s
No Discussion Sections Analyzed	6	7

Table 4.10: Analysis of Discussion Section in Spring of 2005

4.4.2.1 Conclusions for Discussion Section Data

The attempts to improve the discussion section during lab were a guarded success. The number of questions answered increased and time spent per question decreased, relative to results from previous semesters. Both Nu (ν) and Xi (ξ) posed questions in a reasonable amount of time (39 seconds and one minute, 18 seconds, respectively), helped their students to conclude working on problems within a reasonable amount of time (four minutes, 54 seconds and five minutes, 54 seconds, respectively), and had closing presentations of a length such that meaningful questions can be addressed to the presents without taking up too much class time (4 minutes 40 seconds and 5 minutes 54 seconds, respectively). The rewritten laboratory curriculum and weekly TA training is thus judged to be effective at improving the quality and effectiveness of the discussion section.

4.4.3 Lecture Data

The TAs were instructed to avoid using their lecture as an opportunity to question students and rather to leave such activities for the discussion section where they could have interactions with groups. Lectures were supposed to be held to ten minutes; a ROP analysis of their lectures is reported in Table 4.11.

Category	Nu (ν)	Xi (ξ)
Lecture Length	19 m 18 s	12 m 15 s
Talking	43 %	68 %
Writing on whiteboard	32 %	18 %
Demonstrating equipment	12 %	7 %
Total Teacher Centered	**87 %**	**93 %**
Questioning students	7 %	6 %
Answering students questions	4 %	1 %
Students work on board	0 %	0 %
Total Student Centered	**11 %**	**7 %**
Number of Lectures Tracked	6	7

Table 4.11: ROP Analysis of Lecture for TAs Studied in Fall 2005

4.4.3.1 Conclusions from Lecture Data

Analysis of lecture data reveals that TAs were successful at staying on task during lecturing, but were not able to meet the target goal of ten minutes, as Nu's (ν) lectures averaged 19 minutes, 18 seconds and Xi's (ξ) lectures averaged 12 minutes, 15 seconds. This was likely due to their trying to cover too much information and spending insufficient time practicing their lecture before lab.

They did keep their lectures very oriented towards instruction as evidenced by Xi (ξ) spending 87% of his time and Nu (ν) spending 93% of his time directly instructing student. This helped to keep the lectures as short as possible and thus is a desirable behavior. Both TAs did ask occasionally during their lectures if all the students understood what had been presented thus far, and this is where most of the student centered activity comes from. While the lectures are supposed to be short asking students if they understand is not at all inappropriate and reflects a good use of classroom time.

4.4.4 Laboratory Execution

4.4.4.1 Time Motion Studies

In the spring of 2005 the time-motion studies were divided into early (first hour and a half of class) and late (remainder of class) so as to reflect any changes in TA behavior as time passed. It was theorized that TAs might spend more time being idle toward the end of the class and without splitting the observations it was impossible to test for this behavior.

Category	Nu (ν)	Xi (ξ)
Max Engagement (early/late)	(49%/56%)	(50%/16%)
Initiate/Respond	43 % / 57 %	54 % / 46 %
Initiate Time	1 m 23 s	1 m 3 s
Respond Time	1 m 35 s	57 s
No. Time-Motion (early/late)	(6/5)	(7/5)

Table 4.12: Time-Motion Studies for TAs in the Fall of 2004

In Table 4.12 presents the interaction styles of the two TAs. Note that engagement seldom rose above 50% and that interaction time was similar if TA initiated or responded to contact. At the beginning of the semester TA interaction with equipment was tracked, however since instance of interaction was extremely low this protocol was abandoned.

4.4.4.2 Conclusions from Time-Motion Studies

Nu (ν) tended to speak with students fairly constantly throughout the course of the laboratory, remaining engaged in student activities right up until the end of the class (early engagement 49%, late engagement 56%). Xi (ξ), on the other hand, would speak to every group, satisfy himself that their procedure was correct, and then would leave them to do their experiment (early engagement 50%, late engagement 16%). Neither technique is wrong, per se, as long as the TA has adequately engaged and questioned every group. Before firm conclusions can be drawn about the effectiveness of TA interaction other observational tools must be used, such as the IRSQ protocol, as presented in the next section.

4.4.5 TA Response Profile

During the spring of 2005 the IR:SQ protocol was utilized, and the data were divided into observations made while the TA administered the discussion section and observations made while the TA was conducting lab. (In previous semesters the IR:SQ profile was used without noting if data were being drawn from discussion sections or while the lab experiment was in progress.) Table 4.13 is the aggregate data collected for both TAs separated by behavior in discussion section and behavior while the laboratory was being conducted.

Category	Nu (ζ) Discussion	Nu (η) Lab	Xi (ξ) Discussion	Xi (ξ) Class
Initiate : Statement	9 %	38 %	6 %	10 %
Response : Statement	61 %	52 %	56 %	42 %
Total Declarative	**70 %**	**90 %**	**62 %**	**52 %**
Initiate : Question	13 %	5 %	28 %	22 %
Response : Question	16 %	5 %	9 %	25 %
Total Inquisitive	**29 %**	**10 %**	**37 %**	**47 %**

Table 4.13: IR:SQ responses for TA's in Spring of 2005

4.4.5.1 Conclusions from TA Response Profile

Xi's (ξ) response profile was inquisitive 37% of the time during the discussion section and 47% of the time during the lab, whereas Nu's (ν) response profile was inquisitive 29% of the time in the discussion section and only 10% of the time in the lab. Clearly Xi (ξ) had greater skill at Socratically guiding students than Nu (ν), whose data serves as an example of how the rewritten physics lab curricula failed, in many ways, to bring about a laboratory administered by Socratic dialogue. During the lab execution Nu's (ν) interactions centered around answering students' questions, without working to make them answer their own questions.

Apparently TAs require considerable training to conduct labs Socratically, and the training that took place during lab meetings was inadequate to cause them to transition to interacting with students through questioning.

4.4.5.2 TA Dominating Discussions

Informal observations in previous semesters led to some concern that TAs engaged in Socratic interactions with just one group member, as opposed to trying to provoke debate among group members. Thus in 2005, when possible, the observer followed the TA around the room and divided the TAs time up into fifteen second increments when the TA was speaking to students. If the TA spoke without the students saying anything for a fifteen second interval then a T was marked next to the group. If the TA did not speak during the fifteen second interval he or she interacted with the group then an S was marked next to the group. This technique was used to qualify a TA's tendency to dominate a discussion as opposed to foster discussion. The results sorted by TA are as follows:

1. Nu (ν) spent at least 20 % of his time in group discussions speaking to the students and at most 80 % of his time letting students respond or speak among themselves.
2. Xi (ξ) spent at least 22 % of his time in group discussions speaking to the students and at most 78 % of his time letting students respond or speak among themselves.

These observations reveal that the TAs allowed their students to dominate discussions with them, which was presented as a healthy pedagogical tactic during lab meetings. It is regrettable that this technique was adopted for one semester only, and thus it is not possible to compare Nu's (ν) and Xi's (ξ) behavior to those of TAs who were not instructed to try and administer interactions with groups by having students debate each other.

4.4.6 Closing Discussion

In the spring of 2005 TAs were urged to have their students complete their preliminary data analysis in lab and give short presentations of their conclusions at the end of class. The TAs teaching the course were consistent in asking their students to present.

Nu (ν) used closing-class discussions which featured group presentations. These presentations involved an average of 36 % of the groups and lasted for an average of five minutes, 43 seconds. Of the groups that presented, in roughly 50 % only one group member presented and answered questions. This information was gathered from a total of four closing discussions.

Xi (ξ) also used closing-class discussions which featured group presentations. These presentations involved an average of 57 % of the groups and lasted for an average of two minutes, 27 seconds. Of the groups that presented, in roughly 33 % only one group member presented and answered questions. This information was gathered from a total of five closing discussions.

4.4.6.1 Conclusions from Study of Closing Discussion

The closing discussions were a successful part of the rewritten laboratory curriculum in that they were held on a regular basis and students were able to gather sufficient information to make meaningful presentations to their peers. The only major problem detected with the presentations is that a sizable proportion of the groups had at least one nonparticipating member in the presentations. Careful tracking of what students have presented and awarding of points for presentation would encourage all students to present equally and help to ameliorate this problem.

4.5 Analysis of Rewritten Second Semester Laboratory Quizzes

In 2005 the second semester physics lab was assessed with 35 conceptual questions and 12 non-conceptual questions. Quiz results are presented in Table 4.14.

	Nu (ν)	Xi (ξ) section 1	Xi (ξ) section 2
Conceptual	51 ± 5	50 ± 5	59 ± 4
Non-Conceptual	21 ± 6	32 ± 8	54 ± 10

Table 4.14: Aggregate Quiz Scores ± Standard Deviation for Rewritten Second Semester Laboratories

For the rewritten lab, students' performance on conceptual items was considerably higher than their performance on non-conceptual items. This could be indicative of exceptionally good training in conceptual reasoning or exceptionally poor imparting and memorization of basic scientific facts.

4.5.1 Analysis of Quiz Results

Very little further analysis can be done in lieu of this data. The conceptual scores in both classes are close enough to each other that there is no significant difference between the two. The differences in the non-conceptual scores are interesting, however since the non-conceptual items include many categories it is difficult to speculate why this dispersion occurs. It could be due to differences between the populations of students, differences in TA teaching styles, time of day when tests are taken, or some other factor.

The rewritten second semester laboratories included different topics than the prior second semester laboratory, making direct comparison of quiz scores as a method of assessing the effectiveness of classroom reform impossible. It was hoped that the first and third semester laboratories would be rewritten so that a direct comparison could be done between the quiz data collected pre and post rewriting for those two class.

CHAPTER 5

CONCLUSION

5.1 Overview of Chapter

This dissertation began as a study of the effectiveness of the physics laboratory curricula currently employed for physics and astronomy students in the first two years of their undergraduate work. These labs were designed to provide students with an opportunity to explore science under the supervision of an expert. Observational protocols were designed and taken from the work of others in order to assess how effectively TAs were able to administer Socratically driven laboratories and thus how well they were able to realize the ambition of guiding students to discover new facets of science and acquire skills to investigate problems scientifically.

As time passed the focus of the dissertation shifted as qualitative and quantitative observations revealed that TA behavior was divergent from what the authors of the laboratories supposed and it became unclear whether the laboratories were being administered effectively. Observational focus shifted to quantifying TA behavior and comparing how TAs administered the different laboratory sections. These observations eventually led to the conclusion that TAs were not administering their laboratories Socratically and that their behavior was, in general, such that the labs could not be considered inquiry based.

The information that was gathered from the first two semesters of the study were used to inform the rewritten second semester laboratory manual and the new techniques for TA training that were used in that course in the spring of 2005.

In the hopes that innovations will eventually be implemented at this university that ameliorate the issues that the TAs had with administering Socratically driven laboratories this chapter contains a list of recommendations as well as closing thoughts on the study as it was conducted. A separate document entitled *Study of and Recommendations for Inquiry Based Labs in the University of Arizona Physics Department* deals with specific recommendations and notes for each laboratory and the reader is asked to refer to this for specific suggestions.

5.2 General Conclusions

Socratically administered labs are not unknown to the physics education community. Hake, for example [81] reported on labs of his creation that are administered by Socratic dialogue. In his 1992 paper he reported good results with the use of these labs, however he did not report any data on TA behavior of the sort gathered by this study. There is one significant difference between the design of Hake's labs and the revised and rewritten physics laboratory curricula at the University of Arizona and that is that Hake's labs are full of specific instructions for students. Thus in Hake's labs, the lab book itself is written in such a way as to help the students step through the reasoning process while making occasional contact with their TA. The revised physics laboratories in this study have no such analog, they are administered entirely by the TA and the lab manual, by deliberate design, does not contain explicit instructions on how to proceed with the lab.

The main result of this study is that TAs teaching the same lab can have extremely different modes of behavior. Lecture lengths observed ranged from over 40 minutes to under ten minutes. Some TAs studied interacted extensively with groups, walking around the classroom and talking to students, and some sat at the front of the room and answered students' questions when they approached. Close inspection of how TAs were interacting with students indicated that some TAs really did try hard to use questioning to guide their discussions with students, but many did not, preferring to simply provide information directly.

There is a very strong tradition in physics that an instructor's classroom is his/her territory and he/she should be allowed to run it autonomously. This classroom culture may have arisen because teaching physics requires a large amount of content knowledge and the great majority of people who are knowledgeable enough about physics to teach it feel that they do not have time and should not be required to study educational theory and practice. Whatever the reason, it is very difficult for educational initiatives of one sort or another to permeate into the practice of physics teaching.

A significant amount of educational research indicates that not all teaching styles are equally effective. The premise of the revised physics labs was that TAs would be very proactive in leading students to investigate different and interesting aspects of science. The conception that the TAs were well prepared to and interested in managing their classrooms that way could easily be incorrect.

Or, alternately, the model of a Socratically administered lab may simply not work without *highly* qualified TAs, which the physics department does not possess and is not training. The second semester lab manual was rewritten in an effort to reconcile the desire for a Socratically administered lab with the behavior patterns observed in TAs. This attempt was only partially successful, though it did result in TA behavior that was somewhat more question oriented and in much more efficient classroom management.

If the desired outcome of laboratories is to have students participate in discussion sections, complete their analysis, and present results in class, then the TA notes associated with the revised curricula is inadequate and should be revised. If the goal of the laboratory is to have students struggle with contextually difficult experimental and theoretical issues, then the system for preparing TAs to administer their classrooms as it is currently designed is inadequate.

5.3 General Threats to Validity

In the physical sciences simulations are often run and mathematical problems solved by manipulating parameters that somewhat simplify the situation, thus leading to solutions that are generalizable only to situations with similar parameters. It is difficult to simplify the parameters in a social science experiment. Observation protocols are a reliable means of generating accurate information about what is happening in a classroom, however any behavior that an observation protocol has not been specifically designed to measure will not be recorded.

In physical sciences experiments can usually be repeated until a large enough data set to isolate a statistically significant signal is amassed. In social sciences this is often impossible to achieve as environmental conditions beyond the researcher's control can change from year to year and/or an insufficiently large number of subjects are part of the data set to isolate signal from noise. In this study both of these effects are significant.

Social science projects are generally judged in terms of both their external and internal validity. Projects that have strong internal validity show clear and concise relationships between different and relevant variables. Projects are externally valid if the results of the studies can be generalized to different populations of students or different settings.

Several threats to external validity can be identified for this study. The fact that all of the students studied are physics and astronomy majors makes results difficult to generalize to other environments in which labs are taught. Implementation of similar laboratories in other introductory physics classes could yield more generalizable results, however currently there are no plans for any such action.

Also, due to the fact that some of the students were under the age of 18 when they began taking the first semester laboratory, some students were unable to immediately participate. This not only reduced the sample size, but also presented a minor selection problem insofar as some students were not able to immediately participate.

The remainder of this section deals with threats to internal validity, which are more numerous and of greater concern in this situation than the threats to external validity. The project did not include a control group, which would be a TA teaching a course using a traditional lab manual and not teaching Socratically. Comparison of TAs teaching different populations is also not possible as all physics majors are required to take the labs so that they can pass the attendant lecture. The nearest population to the physics majors, the engineering majors, are also required to take labs, however these labs are not written in the same way as the physics labs are. While many of the engineering labs use the same equipment and have the same rough goals as the physics labs they are far more cookbook oriented, and this project only focused on the labs for physics majors, so the engineering labs were never used as a contrasting group.

While many different assessment techniques were used in this study no one particular part of the lab was approached in a thorough enough way that the study could be described as triangular (assessment of one particular form of behavior from multiple perspectives). Variables such as TA lecture length, methods by which TAs respond to questions posed in class, and amount of time TAs spend interacting with students are all assessed via only one observational protocol each due to the fact that only one observer was present in the classroom to administer the protocols.

An additional internal validity threat is that TAs had difficulty understanding some of the material in the labs, which led to confusing and difficult to code behavior. The TA notes (before rewritten by the author) did not cover content knowledge and thus tacitly assumed that the TAs would iron out all difficulties with the concepts in the lab during the lab meeting. If this didn't happen then the TAs were expected to do the experiment themselves so that any errors in their thought processes would become self-evident. However, it is not clear that all TAs engaged in this behavior. Any misconceptions that TAs held before lab began may be transferred to their students or resulted in a confusing or incoherent lab.

5.4 Recommendations

5.4.1 Improved TA Training

Using half of the weekly TA meeting to conduct TA training was potentially useful to TAs, however did not result in the desired interaction profiles. In the spring of 2005, both Nu (ν) and Xi (ξ) exhibited ideal administrative behavior in most measurable ways, however studies of their interaction with students showed that they were still not using Socratic dialogue extensively to interact with students, as the ratio of declarative vs. interrogative statements ranged from 90%/10% to 52%/48%. A TA administering a class through questioning would consistently ask more questions then provide answers, and the data collected in this study clearly shows that this did not happen.

Potentially more practice for TAs could change the way that they administer their classrooms. Videotaping or audiotaping their in-class performance and reviewing it later might also have a positive impact, though it is a time consuming activity. Having a person in a position of perceived power, perhaps a professor, conduct the TA training on Socratic dialogue and wait time might also improve TAs' response to role-playing as a training mechanism.

5.4.2 General Lab Layout

5.4.2.1 Format of Lab

Student comments in interviews were very consistently negative about the format of introducing the lab in the form of a short story. Their aggregate comments indicated that they viewed this less as an opportunity to design their own experiment given an interesting situation and more as a situation in which they should simply come to class and do what the TA said. Thus there was a profound disconnect between the expectations expressed in the laboratory manual and the ways the labs were conducted before the second semester laboratory was rewritten. It is strongly recommended that either the premises of the laboratories be removed, or else they be rewritten so that they are clearer about potential experiments.

The premises of some of the stories are culturally inappropriate for either females or foreign students. Some of the premises of the labs include the student:

- hunting terrorists in Afghanistan for the U.S. Millitary.
- aiding a Cub Scout (a somewhat homophobic and male-dominated group) in winning a contest.
- assisting the alchemist Al-Faquir and will be beheaded if not successful with an experiment.
- having been hired by the NSA to work in Iraq on anti-terrorism issues.

These labs should be rewritten in order to not be exclusionist towards any one particular gender or racist towards any group.

5.4.2.2 Grading of Labs

In general, laboratory TAs grade on a scale of their own design, which could emphasize any element of the course or be completely based on grades earned on laboratory reports. In the past, different lab TAs assigned to teach different lab sections of the same course graded with sufficiently different emphases and severity that considerable student disgruntlement resulted.

It is suggested that a uniform grading scale be established for each laboratory course, possibly including some assignment of points to activities such as giving clear and concise presentations during discussion sections and closing discussions, in order to reflect in the grading the overall philosophy of the course. Grading using rubrics is relatively straightforward and helps to match student performance with a number of different and clearly defined standards.

5.4.2.3 Final Exam

Through 2004, final examinations in physics laboratory courses, when administered, consisted of a series of performance tests (as described in section 2.5.2.1). However, the goals of the revised laboratories include more than acquisition of experimental skills. Students are supposed to learn the fundamentals of group work, presentation and defense of data, and experimental design. To the end of designing an exam to better test these qualities a final exam of dramatically different design was created and used in the rewritten laboratory curricula. This exam may be seen is the associated document *Study of an Recommendations for Inquiry Based Labs in the University of Arizona Physics Department.*

5.4.2.4 Rewrite of Physics Toolkit

The technical manual at the back of the first and second semester laboratory manuals referred to as the “physics toolkit” is in some need of revision. At present it is overly wordy, contains excess information that students do not require, and its sections are not listed in the table of contents of the laboratory manual. Even if the toolkit is not rewritten it is strongly suggested that the first semester laboratory manual contain explicit instructions for students to look in various sections of the toolkit in order to increase the students’ awareness of its existence. A suggested alternative for the toolkit is presented in *Study of and Recommendations for Inquiry Based Labs in the University of Arizona Physics Department.*

5.4.3 Discussion Section

5.4.3.1 Preparation of TAs

It is difficult for TAs to conduct discussion sections when they have never practiced doing so. Timing student discussions, posing questions succinctly, constantly interacting with groups to challenge them to state their assumptions and guiding them when they have difficulties are all necessary components of conducting discussion section.

TAs should be trained to conduct complicated discussion sections of the sort described in sections 2.4.1 and 2.4.2, depending on the needs of the particular class to which they are assigned. Examples of useful training techniques could include role-playing through discussion sections, observing experienced TAs facilitating discussion sections, and coaching by an expert who analyzes the discussion section.

5.4.3.2 Inclusion of Discussion Questions

Until the second semester lab manual was rewritten, TAs were instructed to administer discussion sections but were not given a list of appropriate questions to present to students. This meant that either discussion questions had to be solicited from students (which they were often unprepared to give), TAs had to design their own discussion questions (which were occasionally inappropriately hard or out of sync with the class), or discussion sections were skipped. If the desired goal is a well-run discussion section, then TAs should be provided with discussion questions and the answers to these questions, and the questions should be reprinted in the lab manual to aid TAs in assigning them quickly.

5.4.4 Class Execution

5.4.4.1 Lecture

TAs occasionally demonstrated evidence of fairly serious misconceptions concerning the science that the laboratory was targeting and providing them a sample lecture assists in eliminating these potential misconceptions before class begins. A more serious concern is the fact that TAs occasionally gave lectures that lasted almost an hour, which subtracts dramatically from the laboratory experience and contributes little to student learning. When these labs are further revised, care should be taken to clearly communicate to TAs that in laboratories appropriate lectures are short lectures or, when possible, no lectures at all.

5.4.4.2 Lab Facilitation

TA behavior while administering the labs was less than optimal. Observations established that some TAs interacted with their students almost entirely through answering students' questions, without either posing questions of their own or using questioning to force students to answer questions. Using Socratic dialogue to administer a classroom requires considerable practice and determination. The revised laboratory curricula is written with the tacit assumption that graduate students would be able to administer a classroom this way, however that assumption was incorrect.

Physics graduate students probably can be trained to use Socratic dialogue to administer laboratories, however this would require considerable retraining, in-class observations and critiques, and feedback from a supervisor who is highly involved in training TAs.

It may be much easier to alter TAs' behavior so that they ask more questions by giving them prompting questions in the TA manual. This was done to some extent in the rewritten laboratory manual. If TAs were provided with a list of questions that they could use throughout the lab to stimulate student discussion this could aid the TAs who are not naturally disposed to using Socratic dialogue in engaging students.

5.4.5 Closing Discussions

In theory the TAs have always been asked to have students complete their experiments in class and present their results to each other. In practice this happens much more effectively when TAs are given acceptable closing discussion topics which they can assign to students as class progresses. If this is not done, it is recommended that TAs not be required to orchestrate closing discussions at all.

5.4.6 Uncertainty Analysis

One concern revealed in student interviews and in classroom observations was a lack of an overall strategy to teach the students error analysis. Even if particular TAs (like Theta (θ) or Iota (ι)) took pains to teach their students uncertainty analysis, other TAs did not concern themselves with the matter at all. TAs teaching later labs in the sequence had no way of knowing whether students have been prepared to do proper uncertainty analysis. An interview done with a student who had just finished the entire sequence of revised labs contains a discussion that illuminates this situation well. The student is identified with the letter A, and the interviewer is identified with the letter I.

- A: There should be an emphasis on error analysis, so that...in a lot of the labs where we had to do error analysis a lot of people were lost and didn't know...what exactly to do or what error analysis was all about, or what it meant.

- I: Yeah, it's interesting, because I notice that in (the fourth semester lab) you were often asked to do error analysis. Are you saying that there is no point in the curriculum where error analysis is introduced to you?
- A: I don't think there was a point where it was ever introduced, at all. I remember being asked to do it in (the second semester lab) and I never did. I did do it in (the third semester lab) but it kinda, it wasn't, I was kinda lost and I had to, like, y'know figure things out on my own a lot, I think. Especially in a lab if you're going to be teaching students how to do an experiment you should also teach them error analysis, I think.
- I: At what point do you think that error analysis should be introduced?
- A: Um...first semester. First day of lab.

Although interviews only represent a few selected individuals, here is evidence that at least one student had completed the entire course of labs and had never been introduced to error analysis. She had also never taken the time to systematically learn uncertainty analysis on her own, which would have been the responsible thing to do. Specifically instructing students that they are responsible for learning elements of uncertainty analysis with certain labs might fix this problem.

Teaching uncertainty analysis directly to students might also resolve this issue, but if it is left purely to the discretion of the TA as to when this instruction occurs, then students may end up with gaps in their ability to use uncertainly analysis, a vital skill for any practicing physicist. Assigning TAs a timetable for teaching uncertainty analysis, providing them with material to do so, and revising the uncertainty analysis appendix of the lab to coordinate with the curricular material used on uncertainly analysis would aid physics students in learning uncertainty analysis and would help to ensure that all students learn something of uncertainty analysis.

5.4.7 Quiz Results

The aggregate quiz scores collected over the course of this study, either of the revised or rewritten laboratory curricula, suggest that the laboratories do not have a strong impact on students' understanding of physics. The individual quiz results reprinted in the associated document are suggestive of conceptual difficulties that students have, even relatively late in the semester. Some indicators, such as difficulties that students evidence in analyzing simple circuits by the end of the third semester laboratory, are indicative of the laboratory failing to impart skills which it focuses specifically on providing. The open-ended method of investigation suggested by the laboratory design may actually impede students in learning certain basic skills due to a lack of emphasis on investigating specific topics and a lack of testing and accountability to ensure that these topics have been adequately investigated.

5.4.8 Overall

While the revised physics laboratory curricula is ambitious in many respects it was not written taking into account the character and training of the TAs who teach at the University of Arizona. Currently the laboratories are potentially harmful to students because they are administered so differently than anticipated. A return to a laboratory format where students have the experiment more clearly described in the laboratory manual or considerable extra work in training and preparing TAs would partially or completely mitigate this problem.

APPENDIX A

CONSENT FORM

The following document was given to all students participating in the study. At the beginning of each semester they were provided with the document and instructions to read it carefully and sign and date it if they wished to participate. Students were also verbally reminded at that time that they could withdraw from the study at any time. After the consent forms were signed by the students they were individually signed by the author, photocopied, and the photocopies returned to the students.

Copies of all consent forms are now being kept in a secure location. After this dissertation is completed the consent forms will be shredded and recycled.

Note that the format of the consent form has been changed in order to be LaTeXcompatible. Originally the consent form was formatted such that it fit on two sides of a single sheet of paper.

> ”Subjects Consent Form”
>
> A Study of the Pedagogical Efficacy of the Revised Physics Lab Curricula
>
> I AM BEING ASKED TO READ THE FOLLOWING MATERIAL TO ENSURE THAT I AM INFORMED OF THE NATURE OF THIS RESEARCH STUDY AND OF HOW I WILL PARTICIPATE IN IT, IF I CONSENT TO DO SO. SIGNING THIS FORM WILL INDICATE THAT I HAVE BEEN SO INFORMED AND THAT I GIVE MY CONSENT. FEDERAL REGULATIONS REQUIRE WRITTEN INFORMED CONSENT PRIOR TO PARTICIPATION IN THIS RESEARCH STUDY SO THAT I CAN KNOW THE NATURE AND RISKS OF MY PARTICIPATION AND CAN DECIDE TO PARTICIPATE OR NOT PARTICIPATE IN A FREE AND INFORMED MANNER.
>
> PURPOSE I am being invited to participate voluntarily in the above-titled research project. The purpose of this project is to evaluate the effectiveness of the labs that I am taking in sequence with the attendant physics course.

SELECTION CRITERIA I am being invited to participate because I am signed up for a physics course which uses this laboratory sequence. Approximately 100 subjects will be enrolled in this study in a given semester. If I am under 18 I can participate in this study but my parents must sign this form.

STANDARD TREATMENT(S) If I do not wish to participate in this study then I participate in the lab course as normal. Assessment tools such as quizzes will still be factored into my final lab grade. The results of my grades will not, however, be entered into the study.

PROCEDURE(S) If I agree to participate, I will be asked to consent to the following: 1. Regardless of participation in the study I will be expected to complete anything that will be part of the final grade assessment for the course. 2. I may be asked to grant interviews. 3. Some additional questionnaires may be given which will have no impact on my grade.

RISKS There are no known risks to this study.

BENEFITS There are no direct benefits to my participation in this research project.

CONFIDENTIALITY My confidentiality will be maintained by Mr. James Little. Only he will know what my scores on these tests are for evaluation purposes. (My TA will have access to those scores which they need to assess my final grade). Dr. Ingrid Novodvorsky will also have access to the raw data, but she will not have access to my name.

PARTICIPATION COSTS AND SUBJECT COMPENSATION Students, if they choose to, can allow Mr. Little to conduct 15 to 30 minute interviews. No compensation will be given for these interviews.

CONTACTS I can obtain further information from the principal investigator Mr. James Little, M.S. at (520) 626-8813. If I have questions concerning my rights as a research subject, I may call the Human Subjects Committee office at (520) 626-6721.

AUTHORIZATION BEFORE GIVING MY CONSENT BY SIGNING THIS FORM, THE METHODS, INCONVENIENCES, RISKS, AND BENEFITS

HAVE BEEN EXPLAINED TO ME AND MY QUESTIONS HAVE BEEN ANSWERED. I MAY ASK QUESTIONS AT ANY TIME AND I AM FREE TO WITHDRAW FROM THE PROJECT AT ANY TIME WITHOUT CAUSING BAD FEELINGS. MY PARTICIPATION IN THIS PROJECT MAY BE ENDED BY THE INVESTIGATOR FOR REASONS THAT WOULD BE EXPLAINED. NEW INFORMATION DEVELOPED DURING THE COURSE OF THIS STUDY WHICH MAY AFFECT MY WILLINGNESS TO CONTINUE IN THIS RESEARCH PROJECT WILL BE GIVEN TO ME AS IT BECOMES AVAILABLE. THIS CONSENT FORM WILL BE FILED IN AN AREA DESIGNATED BY THE HUMAN SUBJECTS COMMITTEE WITH ACCESS RESTRICTED TO THE PRINCIPAL INVESTIGATOR, MR. JAMES LITTLE OR AUTHORIZED REPRESENTATIVE OF THE PHYSICS DEPARTMENT. I DO NOT GIVE UP ANY OF MY LEGAL RIGHTS BY SIGNING THIS FORM. A COPY OF THIS SIGNED CONSENT FORM WILL BE GIVEN TO ME.

Subject's Signature ________________________________ Date _____

Parent/Legal Guardian (if you are under 18) ________________ Date _____

INVESTIGATOR'S AFFIDAVIT I have carefully explained to the subject the nature of the above project. I hereby certify that to the best of my knowledge the person who is signing this consent form understands clearly the nature, demands, benefits, and risks involved in his/her participation and his/her signature is legally valid. A medical problem or language or educational barrier has not precluded this understanding.

Signature of Investigator ___________________________ Date _____

]

1/2000

APPENDIX B

HARVARD'S LIST OF LABORATORY EXPERIMENTS

Harvard's list of forty laboratory experiments is presented to illustrate how laboratory tasks and assignments have not changed dramatically in over a century. This list was published in 1886 and is copied Rosen's article *A History of the Physics Laboratory in the American Public High School (to 1910)*, [37, p. 200-201].

1. *Volume Experiments*
    (a) Direct measurement of volume – (Density)
    (b) Displacement
2. *Density*
    (a) Principle of Archimedes
    (b) Density of solid by flotation
    (c) Densimetry
    (d) Density of air (rough).
3. *Pressure*
    (a) Estimation of pressure at different depths.
    (b) Comparison of densities. – Method of balancing columns.
    (c) Mariotte's tube. – (Estimation of atmospheric pressure by law of Boyle and Mariotte).
    (d) Construction and use of barometer.
4. *Mechanics*
    (a) Elasticity.–Stretching of Wire–(Relation of stress to strain.–Flexure).
    (b) Breaking strength of fine wires.
    (c) Coefficient of friction.–(Work)

  - (d) Bent lever.–(Comparison of masses).–Moments of force.
  - (e) Center of Gravity

5. *Dynamics*
   - (a) Law of falling bodies.
   - (b) Velocity of falling bodies.
   - (c) Laws of pendulum.–Influence of length of arc.–Variation of length of pendulum.

6. *Sound*
   - (a) Pitch
   - (b) Comparison of two tuning forks.–Graphical method.
   - (c) Resonance tube.–Wavelengths
   - (d) Velocity of sound.

7. *Heat*
   - (a) Temperature.–Fixed points of thermometer.–Melting and boiling points.
   - (b) Expansion by heat.–Air thermometer
   - (c) Mixing hot and cold water.
   - (d) Specific heat.–Method of mixture.
   - (e) Latent heat of water.
   - (f) Latent heat of steam.
   - (g) Conduction of heat.
   - (h) Loss of heat by convention and radiation–Law of cooling.

8. *Light*
   - (a) Photometry.
   - (b) Law of reflection.–Angle of prism.
   - (c) Refraction.–Principle of spectroscope.–Angle of deviation.

(d) Principle focus of lenses.–(Practical examination of telescope and microscope).

9. *Electricity and Magnetism*

(a) Tracing lines of magnetic force.–Electromagnet

(b) Dipping needle.

(c) Electrical attractions and repulsions.–(Arrangement of series in electropositive order).

(d) Distribution of electrical charge.–Proof plane.(Absence of charge in the interior of a conductor).

(e) Comparison of emf by Ohm's law

(f) Comparison of resistances by Ohm's law.

APPENDIX C

IN-DEPTH ANALYSIS OF POPULATIONS

The issue of student retention is wrapped up in the effect that instruction has on students' attitudes. One of the best ways to determine whether students are "liking" physics is to check whether or not they stay in physics. Using enrollment data, a cohort study was implemented to examine the dropout rate from previous years and compare this to the dropout rate after these labs are instituted.

The data in this appendix represents in depth analysis of the participants in the study. This analysis was not used to make any conclusions vital to the study as a whole, so it is presented in this appendix rather than in the main body of the dissertation. For the sake of brevity the first semester laboratory will occasionally referred to as "first," the second semester laboratory as "second," etc.

C.1 Population of classes

In Table C.1 the number of students in each laboratory section in spring of 2004 and the number of students participating in the study are presented. Note that Epsilon's (ϵ) first class is not included in the study due to its extremely low number of students and the fact that no observer was available for the class, preventing any conclusions informed by qualitative observation. The enrollment data collected from Epsilon's (ϵ) second class is reported here, but not elsewhere. The Tables C.2 and C.3 present the student enrollment by section for fall of 2004 and spring of 2005, respectively.

TA	No. Students Participating	No. Students Total
Phys 152		
Alpha	11	13
Beta	17	19
Gamma	19	20
Summation:	**47**	**52**
Phys 252		
Delta	17	17
Epsilon 1	4	8
Epsilon 2	16	20
Summation:	**37**	**45**

Table C.1: Distribution of Participating and Total Population in Classes Surveyed for Spring of 2004

TA	No. Students Participating	No. Students Total
First Semester Lab		
Theta 1	14	20
Theta 2	13	21
Iota	12	22
Kappa	17	22
Summation:	**56**	**85**
Third Semester Lab		
Zeta	13	22
Eta 1	8	11
Eta 2	10	11
Summation:	**31**	**44**

Table C.2: Distribution of Participating and Total Population in Classes Surveyed for Fall of 2004

TA	No. Students Participating	No. Students Total
Second Semester Lab		
Nu	12	17
Xi 1	9	16
Xi 2	9	10
Summation:	**30**	**43**
Fourth Semester Lab		
Lambda	11	16
Mu	16	22
Summation:	**27**	**38**

Table C.3: Distribution of Participating and Total Population in Classes Surveyed for Spring of 2005

C.2 Student Demographics by Major

These laboratories are coupled to lecture courses designed for physics and astronomy majors, thus the population being tested should be comprised entirely of those two groups excepting for fluctuations due to schedule conflicts and other exceptions. Participation in this study is entirely voluntary, and this is the only segment of this dissertation in which any information on students not agreeing to participate in the study will be presented, and the only information presented here will be division of cohorts by major, gender, and participation.

Each class will be labelled by the year that the student entered the university, assuming that the students in the lab began taking the first semester laboratory the first semester that they arrived at the University of Arizona and continued taking the courses so that they finished the fourth semester laboratory the second semester of their sophomore year. This naming convention does not mean that this approximation is true for all students participating and not participating in the study.

	Fourth	
Major	Participating	Total
Physics	9	10
Astronomy	7	8
Phys & Astro	7	14
Physics & Other	4	4
Astronomy & Other	1	1
Engineering (Physics)	4	4
Engineering (Other)	0	0
Math	1	1
No Major Selected	1	1
Other	2	2
Summation:	**36**	**45**

Table C.4: Cohort of 2003 Sorted by Participation

The cohort of 2003 represented in table C.4 reflects only one semester worth of study, hence in is difficult to make any conclusions about how the composition of the class has changed over time. In the 2003 class 80 % of the students participated in the study, allowing for a reasonable sampling of student understanding.

	Second		Third		Fourth	
Major	Participating	Total	Participating	Total	Participating	Total
Physics	12	13	7	10	4	7
Astronomy	19	19	7	13	9	12
Phys & Astro	2	4	6	6	9	10
Physics & Other	0	0	1	1	0	1
Astronomy & Other	2	2	1	1	1	1
Engineering (Physics)	5	6	2	4	2	3
Engineering (Other)	0	0	2	2	1	1
Math	2	2	1	1	0	0
No Major Selected	4	5	2	2	0	1
Other	1	1	3	4	2	2
Summation:	**47**	**52**	**32**	**44**	**28**	**38**

Table C.5: Cohort of 2004 Sorted by Participation

These tables of data reveal attrition in the physics labs, but offers no supporting evidence that the physics labs had an effect on student retention. In the class that entered in 2004 (see Table C.5) there is an interesting trend representing a tendency for students to become physics/astronomy majors as they spend more time in school. This may be a result of a perception noted among undergraduates to regard the astronomy department as being of extremely high quality, but to perceive that an astronomy degree is not useful without an accompanying physics degree. Participation in the 2004 cohort was very high the first semester (90 %) however fell to 73 % and 74 % for the third and fourth semester laboratories, respectively.

	First		Second	
Major	Participating	Total	Participating	Total
Physics	11	20	10	13
Astronomy	18	26	10	16
Phys & Astro	2	2	1	1
Physics & Other	3	4	0	2
Astronomy & Other	0	0	0	0
Engineering (Physics)	4	5	2	3
Engineering (Other)	1	3	0	0
Math	3	5	1	2
No Major Selected	9	11	2	2
Other	8	9	4	4
Summation:	**59**	**85**	**30**	**43**

Table C.6: Cohort of 2005 Sorted by Participation

The cohort of 2005 depicted in table C.6 was studied for two semesters and took the second semester lab using the lab manual rewritten by the author of this dissertation, so the sampling in the second semester lab is particularly important. Participation in the first semester lab was only 69 %, by far the worst in the study, and by the second semester the participation had only risen to 70 %. Due to the entirely voluntary nature of participation in the project, occasional fluctuations in student participation are to be expected, however the low participation in the 2005 cohort is particularly disappointing since they were the class that received an experimental treatment in the form of a rewritten laboratory manual.

C.3 Student Demographics by Gender

Teasing information about relationships between student majors, their dropout rates, and their genders is a difficult task, especially as many of the students represented in these tables have gone on to change their minds about their majors after their sophomore year of college. Here this data is analyzed to see if there exist different behaviors in terms of majors chosen and dropout rates between peers that can be traced to gender.

	252	
Major	Male	Female
Physics	10 (36%)	0 (0%)
Astronomy	3 (11%)	5 (29%)
Phys & Astro	7 (25%)	7 (41%)
Physics & Other	1 (4%)	3 (18%)
Astronomy & Other	1 (4%)	0 (0%)
Engineering (Physics)	4 (14%)	0 (0%)
Engineering (Other)	0 (0%)	0 (0%)
Math	1 (4%)	0 (0%)
No Major Selected	0 (0%)	1 (6%)
Other	1 (4%)	1 (6%)
Summation:	**28**	**17**

Table C.7: Cohort of 2003 Sorted by Gender

The 2003 cohort represented in table C.7 is, again, difficult to interpret because of the brief period during which they were studied. The class is 37% female during 252 and 70% of the female students are either astronomy or astronomy/physics majors, as opposed to 36% of the males.

	Second		Third		Fourth	
Major	Male	Female	Male	Female	Male	Female
Physics	11 (31%)	2 (13%)	8 (24%)	2 (20%)	5 (20%)	2 (15%)
Astronomy	11 (31%)	8 (50%)	9 (26%)	4 (40%)	9 (36%)	3 (23%)
Phys & Astro	1 (3%)	3 (19%)	2 (6%)	4 (40%)	4 (16%)	6 (46%)
Physics & Other	0 (0%)	0 (0%)	1 (3%)	0 (0%)	1 (4%)	0 (0%)
Astronomy & Other	1 (3%)	1 (6%)	1 (3%)	0 (0%)	1 (4%)	0 (0%)
Engineering (Physics)	5 (14%)	1 (6%)	4 (12%)	0 (0%)	3 (12%)	0 (0%)
Engineering (Other)	0 (0%)	0 (0%)	2 (6%)	0 (0%)	0 (0%)	1 (8%)
Math	2 (6%)	0 (0%)	1 (3%)	0 (0%)	0 (0%)	0 (0%)
No Major Selected	4 (11%)	1 (6%)	2 (6%)	0 (0%)	1 (4%)	0 (0%)
Other	1 (3%)	0 (0%)	4 (12%)	0 (0%)	1 (4%)	1 (8%)
Summation:	**36**	**16**	**34**	**10**	**25**	**13**

Table C.8: Cohort of 2004 Sorted by Gender

In the cohort of 2004 shown in table C.8 the percentage of females in the class are 31% in the second semester lab, 22% in the third semester lab, and back up to 34% by the fourth semester lab. A more telling way to analyze the data may be to look at the relative attrition rates between the genders. In the time that passed between the beginning of 152 and the beginning of 251 the number of female students dropped by 38% while the number of male students dropped by only 6%. Between the beginning of the third semester and the beginning of the fourth semester the number of male students dropped by 34% while the number of female students grew by 30%, accounting for the change in gender balance in the classroom.

The greater bias of female students towards astronomy than male students continues to hold true in this particular cohort. In the second semester class 69 % of the female students were astronomy or physics/astronomy majors, as opposed to 34 % of the male students. In the third and fourth semesters the percentages of astronomy and physics/astronomy to other were, in the third semester, female: 80 % male: 32 % and in the fourth semester the ratios were female: 69 % male: 52 %, reflecting an increase in astronomy majors among the male students.

	First		Second	
Major	Male	Female	Male	Female
Physics	13 (28%)	7 (18%)	8 (32%)	5 (28%)
Astronomy	11 (23%)	15 (39%)	9 (36%)	7 (39%)
Phys & Astro	2 (4%)	0 (0%)	1 (4%)	0 (0%)
Physics & Other	2 (4%)	2 (5%)	2 (8%)	0 (0%)
Astronomy & Other	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Engineering (Physics)	4 (9%)	1 (3%)	2 (8%)	1 (6%)
Engineering (Other)	2 (4%)	1 (3%)	0 (0%)	0 (0%)
Math	3 (6%)	2 (5%)	2 (8%)	0 (0%)
No Major Selected	6 (13%)	5 (13%)	1 (4%)	1 (6%)
Other	4 (9%)	5 (13%)	0 (0%)	4 (22%)
Summation:	**47**	**38**	**25**	**18**

Table C.9: Cohort of 2005 Sorted by Gender

The cohort of 2005 as represented in table C.9 was not studied for more than two semesters, and so consequently it is difficult to develop any firm conclusions about the study. During the first semester lab the student body was 45 % female, and by the second semester the class was 42 % female. The attrition rates were not significantly different amongst male and female students.

C.4 Longitudinal Study

In this section the enrollment data over the last ten years is considered. If the labs have a demonstratively positive effect on student enrollment then retention should improve for the years associated with the classes of 2002, 2003, 2004, and 2005, when the revised labs were implemented. Of course an improvement of retention of students during those years may not be due to laboratory reform at all, but to some other factor entirely.

The physics department only retains records of students who stay in a course until the very end and receive a letter grade, thus the data represented in Table C.10 does not represent attrition that occurs during the progress of a course, simply the attrition (and occasional addition) of students from semester to semester.

Note also that the data in Table C.10 represents students finishing the course, whereas the results reported in the previous three sections report students beginning the course. There will hence be disagreement between the two sets of tables due to students dropping out. The revisions to the laboratory curricula were enacted in 2001, and so only 2001-2003 are representative of attrition rates with the new curricula in place.

	First		Second		Third		Fourth	
	Male	Female	Male	Female	Male	Female	Male	Female
1995	26	8	24	7	17	0	14	0
1996	29	5	18	1	17	2	13	1
1997	22	10	16	7	14	6	12	5
1998	28	19	21	18	15	11	10	10
1999	27	15	25	15	17	14	17	10
2000	35	14	22	10	15	8	20	9
2001	40	17	23	13	20	7	17	7
2002	50	27	33	20	25	17	24	15
2003	43	22	33	13	29	11	17	11

Table C.10: Study of Enrollment from 1994 to 2003

The statistical data in Table C.10 can be analyzed in a number of ways, however for the purpose of this study if the revised physics laboratory curricula is, in fact, more effective at attracting students to the revised laboratory course sequence and in retaining the students that enter then the attrition rates for both male and female students from the first semester lab to the fourth semester lab will drop in the years 2001, 2002, 2003. This data is analyzed in Table C.11 and, as it does not reflect any significant lowering in attrition rate, it seems unlikely that the new laboratory design is having a positive effect on student enrollment.

	M	F	Total
1995	46 %	100 %	58 %
1996	55 %	80 %	58 %
1997	45 %	50 %	47 %
1998	65 %	47 %	57 %
1999	37 %	33 %	36 %
2000	43 %	35 %	41 %
2001	58 %	59 %	58 %
2002	52 %	44 %	49 %
2003	60 %	50 %	57 %

Table C.11: Attrition from first semester to fourth semester from 1995 to 2003

REFERENCES

[1] J. Piaget. *The Construction of Reality in the Child.* Basic Books, New York, 1954.

[2] D. Phillips and D. Phillips. *Structures of Thinking: Concrete Operations.* Kendall/Hunt, 1995.

[3] J. Piaget and B. Inhelder. *The Psychology of the Child.* Basic Books, 1969.

[4] P. Sutherland. *Cognitive Development Today: Piaget and his Critics.* Paul Chapman Publishing Ltd., 1992.

[5] J. W. McKinnon and J. W. Renner. Are Colleges Concerned with Intellectual Development? *American Journal of Physics*, 39:1047–1052, 1971.

[6] L. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes.* Harvard University Press, 1978.

[7] M. L. Bigge and S. S. Shermis. *Learning Theories for Teachers.* Longman, 1999.

[8] L. Vygotsky. *Thought and Language.* MIT Press, 1986.

[9] C. Lerch. The Unfossilization of Concepts in Adult Learning. In *Report CE 085 086*, 2002. (ERIC document ED 477 503).

[10] R. Driver, H. Asoko, J. Leach, E. Mortimer, and P. Scott. Constructing Scientific Knowledge in the Classroom. *Educational Researcher*, 23:5–12, 1994.

[11] S. Carey. *Conceptual Change in Childhood.* MIT Press, 1985.

[12] P. Carmicael, R. Driver, B. Holding, I. Phillips, D. Twigger, and M. Watts. *Research on Students' Conceptions in Science: A Bibliography.* Centre for Studies in Science and Mathematics Education, University of Leeds, 1990.

[13] H. Pfundt and R. Duit. *Bibliography: Students' Alternative Frameworks and Science Education.* IPN, 1985.

[14] J. Piaget. *Genetic Epistemology.* Columbia University Press, 1970.

[15] D. Edwards and N. Mercer. *Common Knowledge.* Methuen, 1987.

[16] J. Lemke. *Talking Science.* Ablex, 1990.

[17] B. Rogoff and J. Lave. *Everyday Cognition: Its Development in Social Context.* Harvard University Press, 1984.

[18] K. Johnston and R. Driver. *A Constructivist Approach to the Teaching of the Particulate Theory of Matter: A Report on a Scheme in Action.* Centre for Studies in Science and Mathematics Education, University of Leeds, United Kingdom, 1990.

[19] P. Scott. *Overtures and Obstacles: Teaching and Learning about Air Pressure in a High School Classroom.* Misconceptions Trust, 1993.

[20] P. Scott, H. Asoko, R. Driver, and J. Emberton. *Working from Children's Ideas: An Analysis of Constructivist Teaching in the Context of a Chemistry Topic.* Falmer, 1994.

[21] D. Hodson and J. Hodson. From Constructivisim to Social Constructivism: A Vygotskian Perspective on Teaching and Learning Science. *School Science Review*, 79:33–41, 1998.

[22] I. A. Halloun and D. Hestenes. The Initial Knowledge State of College Physics Students. *American Journal of Physics*, 53:1043–1055, 1985.

[23] D. Hestenes, M. Wells, and G. Swackhamer. Force Concept Inventory. *The Physics Teacher*, 30:141 – 158, 1992.

[24] D. Hestenes and M. Wells. *The Physics Teacher*, 30:159–166, 1992.

[25] R. R. Hake. Interactive-Engagement versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses. *American Journal of Physics*, 66:64–74, 1998.

[26] C. Henderson. Common Concerns About the Force Concept Inventory. *The Physics Teacher*, 40:542 – 546.

[27] V. P. Coletta and J. A. Phillips. Interpreting FCI Scores: Normalized Gain, Pre-instruction Scores, and Scientific Reasoning Abilitiy. *American Journal of Physics*, 73, 2005.

[28] D. Hammer. More than Misconceptions: Multiple Perspectives on Student Knowledge and Reasoning, and an Appropriate Role for Education Research. *American Journal of Physics*, 64:1316–1324, 1996.

[29] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. V. Heuvelen. Surveying Students' Conceptual Knowledge of Electricity and Magnetism. *American Journal of Physics Supplement on Physics Education Research*, 69:S12–S23, 2001.

[30] R. N. Steinberg and M. S. Sabella. Performance on Multiple-Choice Diagnostics and Complementary Exam Problems. *The Physics Teacher*, 35:150–155, 1997.

[31] D. Huffman and P. Heller. What Does the Force Concept Inventory Actually Measure? *The Physics Teacher*, 33:138–143, 1995.

[32] D. Hestenes and I. Halloun. Interpreting the Force Concept Inventory: A Response to March 1995 Critique by Huffman and Heller. *The Physics Teacher*, 33:502–506, 1995.

[33] P. Heller and D. Huffman. Interpreting the Force Concept Inventory. *The Physics Teacher*, 33:503–511, 1995.

[34] D. S. Domin. A Review of Laboratory Instruction Styles. *Journal of Chemical Education*, 76:543–547, 1999.

[35] C.L. Henshaw. Do Students Find History Interesting in Physical Science Courses? *American Journal of Physics*, 18:373–377, 1950.

[36] Various. *The Harvard Report, General Education in Free Society.* Harvard University Press, 1945.

[37] S. Rosen. A History of the Physics Laboratory in the American Public High School (to 1910). *American Journal of Physics*, 22:194–204, 1954.

[38] H. Spenser. *Education, Intellectual, Moral, and Physical.* Werner, 1860.

[39] M. Phillips. Early History of Physics Laboratories for Students at the College Level. *American Journal of Physics*, 49:522–527, 1981.

[40] *National Educational Association Conference Proceedings*, 1903.

[41] J.R. Prescott and C. D. Anger. Removing the "Cook Book" from Freshman Physics Laboratories. *American Journal of Physics*, 39:271–274, 1970.

[42] F. Rief and M. St. John. Teaching Physicists' Thinking Skills in the Laboratory. *American Journal of Physics*, 47:950–957, 1979.

[43] A. Arons. *Teaching Introductory Physics.* John Wiley & Sons, 1997.

[44] A. Arons. Development of Energy Concepts in Introductory Physics Courses. *American Journal of Physics*, 67:1063–1067, 1999.

[45] H. Kruglak. The Measurement of Laboratory Achievement. *American Journal of Physics*, 22:442–451, 1954.

[46] H. Kruglak. Some Behavior Objectives for Laboratory Instruction. *American Journal of Physics*, 19:223–225, 1951.

[47] H. Kruglak. The Measurement of Laboratory Achievement, Part II. Paper-Pencil Laboratory Achievement Tests. *American Journal of Physics*, 22:452–462, 1954.

[48] H. Kruglak. Measure of Laboratory Achievement, Part III. Paper-Pencil Analogs of Laboratory Performance Tests. *American Journal of Physics*, 23:82–87, 1955.

[49] H. Kruglak. Current Teaching Practices and Problems in the General Physics Laboratory. *American Journal of Physics*, 28:791–793, 1960.

[50] H. Kruglak. Experimental Outcomes of Laboratory Instruction in Elementary College Physics. *American Journal of Physics*, 20:136–141, 1952.

[51] H. Kruglak. Achievement of Physics Students With and Without Laboratory Work. *American Journal of Physics*, 21:14–16, 1953.

[52] S. Brown. Do College Students Benefit from High School Laboratory Courses? *American Journal of Physics*, 26:334–337, 1958.

[53] S. Brown. Elementary Laboratory Instruction in England. *American Journal of Physics*, 23:27–30, 1955.

[54] S. Brown. Elementary Physics Laboratory Instruction at the Massachusetts Institute of Technology. *American Journal of Physics*, 25:334–337, 1957.

[55] D. D. Long. The Influence of Physics Laboratories on Student Performance in a Lecture Course. *American Journal of Physics*, 54:122–125, 1986.

[56] J. I. Shonle. A Progress Report on Open-Ended Laboratories. *American Journal of Physics*, 38:450–456, 1970.

[57] M. P. Freedman. Relationship among Laboratory Instruction, Attitude toward Science, and Achievement in Science Knowledge. *Journal of Research in Science Teaching*, 34:343–357, 1997.

[58] W. S. Toothacker. A Critical Look at Introductory Laboratory Instruction. *American Journal of Physics*, 51:516–520, 1983.

[59] R. Lock. Gender and Practical Skill Performance in Science, 1992.

[60] J. A. Shymanksy, L. V. Hedges, and G. Woodworth. A Reassessment of the Effects of Inquiry-Based Science Curricula of the 60's on Student Performance. *Journal of Research in Science Teaching*, 27:127–144, 1990.

[61] M. Weinburgh. Gender Differences in Student Attitude Toward Science: A Meta-Analysis of the Literature from 1970-1991. *Journal of Research in Science Teaching*, 32:387–398, 1990.

[62] M. C. Linn, T. De Benedictis, A Harris, and E. Stage. Gender Differences in National Assessment of Educational Progress Science Items: What Does "I Don't Know" Really Mean? *Journal of Research in Science Teaching*, 24:257–278, 1987.

[63] M. P. Freedman. The Influence of Laboratory Instruction on Science Achievement and Attitude Toward Science across Gender Differences. *Journal of Women and Minorities in Science and Engineering*, 8:191–200, 2002.

[64] B. Thacker, E. Kim, K. Trefz, and S. M. Lea. Comparing Problem Solving Performance of Physics Students in Inquiry-Based and Traditional Introductory Physcis Courses. *American Journal of Physics*, 62:627–632, 1994.

[65] D. E. Meltzer and K. Manivannan. Transforming the Lecture Hall: The Fully Interactive Physics Lecture. *American Journal of Physics*, 70:639–654, 2002.

[66] P. Heller, R. Keith, and S. Anderson. Teaching Problem Solving Through Cooperative Grouping. Part 1: Group Versus Individual Problem Solving. *American Journal of Physics*, 60:627– 636, 1992.

[67] P. Heller, R. Keith, and S. Anderson. Teaching Problem Solving Through Cooperative Grouping. Part 2: Designing Problems and Structuring Groups. *American Journal of Physics*, 60:637 – 644, 1992.

[68] E Mazur. *Peer Instruction: A Users Manual.* Prentice Hall, 1997.

[69] M. B. Rowe. Wait-Time and Rewards as Instructional Variables, Their Influence on Language, Logic, and Fate Control: Part One-Wait-time. *Journal of Research in Science Teaching*, 11:81–94, 1974.

[70] M. B. Rove. Wait Time: Slowing Down May Be a Way of Speeding Up. *American Educator*, 11:38–43, 1987.

[71] O. K. Duell, D. J. Lynch, R. Ellsworth, and C. A. Moore. Wait-time in College Classes Taken by Education Majors. *Research in Higher Education*, 33:483–495, 1992.

[72] O. K. Duell. Extended Wait Time and University Student Achievement. *American Educational Research Journal*, 31:397–414, 1994.

[73] K. G. Tobin. The Effect of an Extended Teacher Wait-Time on Science Achievement. *Journal of Reasearch in Science Teaching*, 23:191–200, 1980.

[74] T. H. Schram. *Conceptualizing and Proposing Qualitative Research.* Merrill Prentice Hall, 2003.

[75] M. B. Miles and A. M. Huberman. *Qualitatitive Data Analysis.* SAGE Publications, Inc, 1984.

[76] M. Denscombe. *The Good Research Guide.* Open University Press, 1988.

[77] M. R. Abraham and D. M. Schlitt. Verbal Interaction: A Means for Self Evaluation. *School Science and Mathematics*, 73:677–685, 1973.

[78] E. Leonard Jossem. Resource Letter EPGA-1: The Education of Physics Graduate Students. *American Journal of Physics*, 68:502–512, 2000.

[79] The Teaching Assistant Oversight Committee. *What's Working in TA Training 2004.* Center for Teaching and Learning, Stanford University, 2004.

[80] J. Gilreath and T. Slater. Training Graduate Teaching Assistants to be Better Undergraduate Physics Educators. *Physics Education*, 29:200–203, 1994.

[81] R. R. Hake. Socratic Pedagogy in the Introductory Physics Laboratory. *The Physics Teacher*, 30:546–552, 1992.