ABSTRACT

Title of Dissertation: STUDENTS' UNDERSTANDING OF MEASUREMENT AND UNCERTAINTY IN THE PHYSICS LABORATORY: SOCIAL CONSTRUCTION, UNDERLYING CONCEPTS, AND QUANTITATIVE ANALYSIS

Rebecca Faith Lippmann, Doctor of Philosophy, 2003

Dissertation directed by: Professor Edward F. Redish
Department of Physics

In the physical sciences and other fields, conclusions are made from experimental data. To succeed in such fields, people must know how to gather, analyze, and draw conclusions from data: not just following steps, but understanding the concepts of measurement and uncertainty. We design the *Scientific Community Laboratory (SCL)* to teach students to utilize their everyday skills of argument and decision-making for data gathering and analysis. We then develop research tools for studying students' understanding of measurement and uncertainty and use these tools to investigate students in the traditional laboratory and in the SCL.

For students to apply their everyday skills of argument and decision-making, they must be in a state of mind (a *frame*) where they consider these skills productive. The laboratory design should create an environment which encourages such a frame. We determine student's frames through information reported by students in interviews and surveys and through analyzing students' behavior. We find that the time students spend sense-making in the SCL is five times more than in traditional labs. Students in both labs frequently evaluate their level of understanding but only in the SCL does that evaluation cause a change to more productive behavior.

We analyze lab videotapes to determine underlying concepts commonly used by students when gathering and analyzing data. Our final goal is for students to use these concepts to analyze data in an appropriate manner. We develop a multiple-choice survey which asks students to analyze data from a hypothetical lab context. With this survey we find more students using range to compare data sets after the SCL (from 12% before to 43% after).

For students to understand measurement and uncertainty, we argue that the laboratory must be designed to encourage students to be in a frame where they view resources used to argue and evaluate as appropriate, engage in productive behavior and monitor their behavior, use productive resources to build an understanding of the underlying concepts, and use those concepts to analyze data. We make use of interviews, surveys, and video data to study each of these requirements and to evaluate the SCL curriculum.

STUDENTS' UNDERSTANDING OF MEASUREMENT AND UNCERTAINTY IN THE PHYSICS LABORATORY: SOCIAL CONSTRUCTION, UNDERLYING CONCEPTS, AND QUANTITATIVE ANALYSIS

by
Rebecca Faith Lippmann

Dissertation submitted to the Faulty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
2003

Advisory Committee:
Professor Edward F. Redish, Chair
Professor Spencer A. Benson
Professor David M. Hammer
Professor Wolfgang Losert
Professor Frederick C. Wellstood

ACKNOWLEDGEMENTS

Thank you to my parents, Carolyn and Richard Lippmann, for giving me faith;
to my advisor, E. F. Redish, cat-herder and frame-negotiator extraordinaire;
to Peter, my rock, for his unfailing encouragement;
and to the myriads of others, thank you, and remember: eschew surplusage.

This material is based upon work supported by the National Science Foundation under Grant Nos. REC-0087519 and DUE-9652877. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

> The fear of the LORD is the beginning of wisdom,
> and knowledge of the Holy One is understanding.
> Proverbs 9:10

PREFACE

"You're here because you want to know something. How well you know it you can't explain. You've felt it your entire life; that there's some way you can draw conclusions from the world; you don't how, but it's there, like a splinter in your mind, driving you mad. It is this feeling that has brought you to me. Do you know what I'm talking about?"

"The Measurement?"

"Do you want to know what it is? The Measurement is everywhere, it is all around us. Even in this very room. You use it when you look out your window or when you microwave popcorn. You use it when you go to work, when you go to church, when you pay your taxes; it is the data you must analyze to get to the truth."

"What truth?"

"That you are a slave, like everyone else you were born into bondage: experiment or theory; born into a prison that you can smell, taste, and touch or a prison for your mind. Unfortunately, no one can be told what the Measurement is. You have to experience it for yourself."

"How?"

"You take the theory pill and the story ends. You wake in your bed and you believe whatever you want to believe. You take the experiment pill and you stay in the real world and I show you how well you know your data."

Adapted from Wachowski, L. and A. (1999).
The Matrix [Film Script]. Warner Bros.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction and Overview

Introduction

In every physics laboratory students take measurements and must analyze their data to make a conclusion. Teaching uncertainty analysis is a goal of many science laboratories[1], yet professors of all levels admit to being frustrated at students' lack of ability and understanding in these areas. According to one of the two popular texts on this subject, "the analysis of uncertainties, or 'errors', is a vital part of any scientific experiment…but I have found that it is often the most abused and neglected part." (Taylor, 1997, p. xv) People in their everyday lives would benefit from being able to design investigations, and interpret and evaluate experiments and results (Duggan and Gott, 2002). This is even more true for those planning to perform biology research or interpret the results of a medical test, as are more than 70% of the students in the University of Maryland's algebra-based introductory physics class.

Unfortunately, there has been little research on students' understanding of measurement and even less research on students' understanding of uncertainty. Existing research focuses on identifying student difficulties with interpreting single and multiple measurements. (See Sere et al., 1993; Lubben and Millar, 1996; Leach, 1998; and Buffler et al., 2001.) In order to teach measurement we must know how students understand measurement. What must students understand in order to design experiments and interpret data? How do students build an understanding of measurement and uncertainty in a social laboratory setting? Is the design of a lab helping this process? This study attempts to answer these questions for health-science students in the algebra-based physics laboratory.

In the remainder of this chapter, we introduce the ideas critical to this dissertation. We start by defining necessary terms and the main goals, and then describe the theoretical framework (all of which will be revisited in more detail in the appropriate chapters). We then use these ideas to present a diorama of the dissertation.

Measurement

Measurement is the act of quantifying a particular attribute of a system. It is accomplished by interacting two systems: the system we wish to measure, and the instrument we use to measure. The instrument must have a certain property that allows one to quantify the interaction, but also must change the measured system as little as possible. For example, a thermometer may be filled with alcohol which changes volume depending on its temperature. Thus when it comes into thermal equilibrium with a system, the volume of the alcohol indicates the temperature of the two systems. The odometer on a car uses a cable and a series of carefully calibrated worm gears to translate from number of turns of the drive shaft to linear distance traveled.

All this seems straightforward, so why does measurement need to be studied? Suppose a student were investigating whether the heat capacity of corn oil was different from that of water. He might measure out 300 g of hot water and room temperature corn oil and mix them. If the temperature of the mixed liquids is the average of the individual liquids then they have similar heat capacities. The concept of measurement comes into play when the student must decide how to

---

[1] The goal "for the student to learn to handle experimental errors" was ranked very important by all 409 teachers surveyed by Welzel et al. (1998)

measure the temperature and with what. A candy thermometer would be a poor choice: built to withstand boiling sugar and the hazards of kitchen use, it is quite large compared to the samples being measured and when equilibrating with the samples would change their temperature considerably. Suppose the student chooses a small probe he connects to a computer. Should he use one probe for all the measurements or a probe in each liquid?

Suppose an engineer has designed a new go-kart and wants to test its gas mileage. She builds a prototype and plans to measure the distance it can travel on one tank of gas. Should she drive it on a track where the distance is already marked, or attach an odometer? She must choose which is better: the quality of the distance markings or the odometer reading. Will the drag the odometer introduces matter? Both examples require a basic understanding of measurement. Such questions are integral to the design of any experiment.

Uncertainty

Roughly speaking, uncertainty is how well one knows one's data or results. According to Taylor (1997), the terms error and uncertainty are interchangeable[1]; however, error has a conflicting common use, so we choose the term uncertainty. Uncertainty analysis involves propagating uncertainty, comparing sets of data, choosing how to report data, and other related activities. An understanding of measurement and uncertainty are both necessary to design an experiment and interpret results – in any subject and any context (including physics or biology in the student laboratory or the doctor's office).

For instance, suppose a nurse is measuring a patient's blood pressure. He first inflates the cuff so no pulse can be heard in the artery. He then lets out the pressure until he hears a pulse begin (this is the systolic pressure). He continues reducing the pressure until the sound disappears, having reached the diastolic pressure (the blood pressure when the heart is relaxed). He measures 133 mmHg systolic and 87 mmHg diastolic (reported as 133/87 mmHg). Suppose the healthy cutoff is 130/85 mmHg. Should something be done?

One source of uncertainty in this measurement comes from the speed at which the pressure is reduced – this influences when the cuff pressure hits a pulse. Manuals (Bickley, 1999) recommend reducing the pressure at a rate of 3 mmHg/sec, which would result in an uncertainty of 3 mmHg (for a pulse of 60 beats/min). The act of measuring can also affect the patient's blood pressure – both because the patient is nervous, and because of the constriction of the cuff. Even having the patient's legs hanging or resting flat on the floor can cause a difference in the measurement. The nurse must use not only the mean value, but also information about the variability of the measurements to make a diagnosis. The test does not result in a single value, but rather a set of numbers, a range that must be interpreted.

Most measurements result in a range of values, and analyzing those results requires an understanding of the range and its causes. In the above example, the range could be caused by changes in the person's blood pressure (from their position, mental state, level of activity) or by the measurement method or instrument. The same could be said for white blood cell count, or many other biological measurements. The result's interpretation depends not only on the mean value, not only on the spread of values, but also on the cause of the spread.

Goals

Traditionally, the lab can have many purposes, including for the student to link theory and practice and learn experimental skills. Students claim the main goal of traditional "cookbook"

---

[1] Bevington and Robinson (1992) use error to mean specifically the difference between one's data and the accepted value, so that uncertainty leads to error in one's data.

laboratories is for them to learn physics concepts better by having participated in a demonstration of the concept[1], and many physics professors would agree. We further discuss common goals in Chapter 2. Our goal is for students to be able to use a basic understanding of measurement and uncertainty to design, run, analyze, and evaluate their own experiment and other scientific evidence. While making progress toward this goal, there are several necessary sub-goals. To teach measurement and uncertainty, we need to know students' initial state (what they already know), final state (what they need to know) and have a model of how cognitive change occurs. To study our students' understanding, we need data of students struggling with measurement and uncertainty issues and research methods for analyzing that data. To get useful data, we need a curriculum that requires students to think through these issues, the development of which is informed by our theory of student thinking on this topic. (See Figure 1–1.) This cycle of teach, evaluate data, and build a theory will be repeated throughout this dissertation with smaller, more specific ideas.

Theoretical Framework

In the field of physics, we often use different theories depending on the context and upon which is easiest to apply and will answer the appropriate question. For instance, when determining if a thrown basketball will make it through a hoop, one would use kinematics from classical mechanics. It would be foolish to attempt to use quantum mechanics to solve for the wave function of every particle in the ball. However, one would use quantum mechanics to determine where an electron may hit a screen after passing through a slit. When solving problems about the behavior of light, sometimes using a ray diagram is more productive and sometimes using wave superposition is more productive. It is not that the theories contradict each other; they are different methods for interpreting and solving the same problem. Theories on student thinking are roughly analogous. Depending on the context and the population, one theory may be more productive for interpreting the problem and posing a solution. In the following section, we give a brief explanation of some different theoretical frameworks for analyzing student thinking. A more complete account is given in Chapter 2. Later we will use these frameworks to describe the main ideas and goals of this project.

---

[1] See student's responses to the traditional laboratory in Chapter 4.

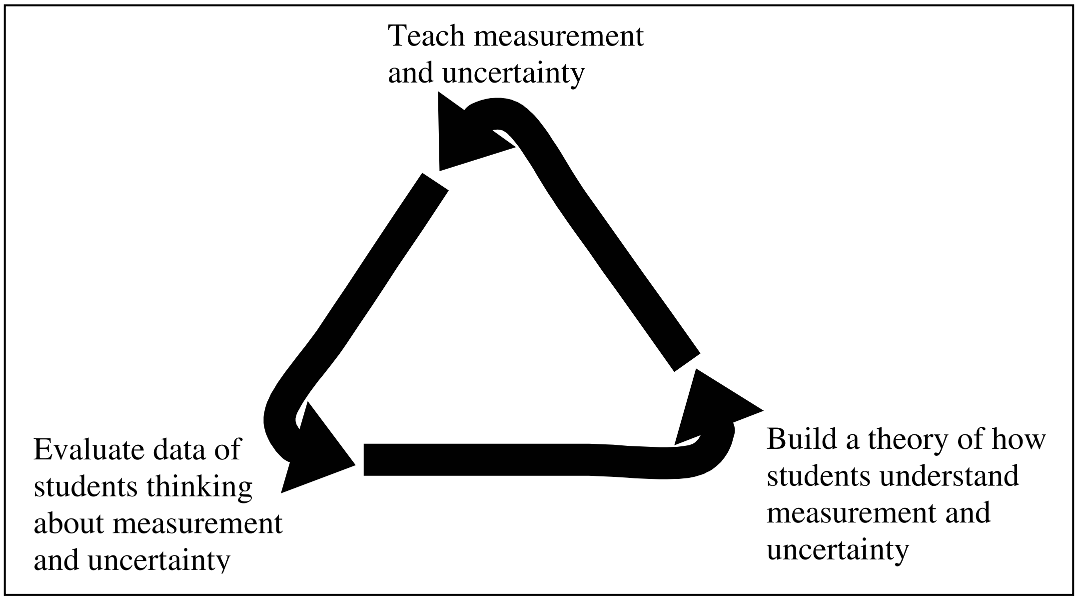


Figure 1–1

The goals of this dissertation and how they interact

Misconceptions

The misconceptions framework (also called alternative conceptions or preconceptions) claims that students have an underlying knowledge structure that they reliably apply to several different contexts. Misconceptions have the following properties: (Hammer, 1996 and Smith et al., 1993)

- They are widespread among (20% or more) students
- They are strongly held and resistant to change
- They interfere with student's understanding and learning
- They must be replaced

One example (McCloskey et al, 1980) is the impetus misconception: a moving object carries with it some impetus (called "force", "momentum", or "energy" by the students) which dies away on its own accord. Students using this misconception would expect a ball shot out of a horizontal circular tube lying on a table to continue following a curved path which would straighten out with time. Students who draw a force from the hand on a ball tossed upward are also applying this misconception (McCloskey, 1983). The goal of instruction is to eliminate the misconception (by showing students it is wrong) and replace it with the correct conception.

Resources

In contrast to a misconception, a resource is more fundamental and abstract, and is used to build an idea in the moment. For the case of the ball thrown upward, a student who draws the force of the hand as acting throughout the ball's path may be applying the general idea of a *maintaining agency* to the situation. The general idea, *maintaining agency*, is the resource, and when the idea is applied to the force of the hand, we call that a *facet*.[1] A resource is built from everyday experience and useful in many different contexts. For example, *maintaining agency* could be used in the following situations: a constant push is needed to move a box along the floor, unceasing nagging is needed to keep a child behaving, and a continual flow of money is needed to keep a judge bribed. A resource is not itself right or wrong, but it can be applied in a context where it is productive or unproductive.

---

[1] The term facet is adopted from Minstrell (1992).

In this example, an instructor would want to help the student apply the resource *actuating agency* and conclude that the force of the hand starts the ball's motion.

How does one know whether to use a misconceptions framework or a resources framework? Returning to our previous analogy using physics, the boundary between classical and quantum is relatively well defined. However, the ability to know which model of light (ray optics, wave superposition, or photons) to apply in solving a problem is built mainly through experience. Physics students learn to use ray diagrams to solve a problem with an object, a lens and an image. This model will not work for solving problems on the photoelectric effect. With some questions one can only try to solve the same problem using the different techniques and observe the outcome. In the same way, whether to use misconceptions or resources is an empirical question. Suppose it is clear that students are applying a more abstract idea on the spot or that they are using a robust, pre-compiled knowledge structure. Even if the data points to misconceptions, a misconception could be a set of resources reliably cued under certain circumstances. In some cases, it might be useful to know the underlying structure of a misconception. In other cases the underlying structure might be so complicated that such an attempt would be useless. One framework may be more productive for a certain context and population, and a researcher must find which has a greater value[1].

We have described two theories useful for interpreting reasoning about physics concepts. There exist other theories (Hammer, 1996) but these two are relatively common in the literature and are used throughout this dissertation. They also can be applied to other areas in addition to physics conceptual knowledge, such as epistemology.

Epistemology

Epistemology deals with ideas about knowledge and knowing, for example where knowledge originates and how to tell if it is valid. A student's epistemology affects the methods a student uses to learn and how well they learn (Hofer, 2001). For example, Tsai (1998) found that students with empiricist beliefs (that scientific knowledge emerges infallibly from objective data) were more likely to learn by memorization than students with constructivist beliefs (scientific knowledge is tentative and continually being invented). Measurement and uncertainty are largely epistemological issues, as they deal with where knowledge comes from and how reliable it is. As in the previous section, one must determine whether a misconception framework or a resources framework is more productive for a question, context, and population. Research has been done claiming that students have misconceptions about epistemology (Schommer, 1990) and that students are applying epistemological resources (Hammer and Elby, 2001). The choice of one or another of these theories will imply different data interpretation and different instructional strategies (Hammer and Elby, 2003).

Frames

A frame is a state of mind related to the larger context and helps determine what actions the student will take. Suppose that a hungry backpacker has just finished a two week trip, enters a diner, and orders a meal. When her food arrives, instead of finding a comfortable rock to sit on, eating with her hands, and wiping out the dish with sand or a leaf, she picks up a knife and fork and eats in a far different manner from when she was on the trail. How does she decide which manner of dining to use? She could check the behavior of other people in the area, the availability of a knife and fork, or other environmental cues. She is able to decode all of this information to determine what behavior is appropriate. Different types of dining establishments have certain expected verbal behavior (drunken cheering for a certain sports team or quiet conversation), food

---

[1] See Scherr (2002) for an example of a comparison between the application of misconceptions or resources.

delivery method (get it when your number is called, or have it brought to the table), trash removal method, payment method, and other expected behavior associated with it. The restaurant environment encourages the hiker to use a different method of eating than the side-of-trail environment.

A frame may be roughly defined as what a person thinks they are doing in a given situation[1], and carries along with it readily observable behavior such as speech patterns, vocabulary, body language, and other accepted behavior. A hiker eating along a trail is in a frame where she wants to get the food inside of her with the least effort possible. A hiker in a restaurant is in a frame where she is more considerate of other restaurant patrons and the cleaning staff, so she eats neatly with a knife and fork.

As relational beings, people are constantly negotiating through different frames: a professor might argue a scientific theory with a colleague, and then be accosted by a student pleading for a better grade, all the while placating his child who has accompanied him to work. The professor could say “if you behave we’ll go get ice-cream in ten minutes” to his child, but great confusion would result if that were said to a colleague. The professor also would likely use different methods to convince the colleague of a theory and to convince the student that their grade is fair and deserved. In the latter, it would be appropriate to refer to his authority as a professor, “that was my decision”; however, an appeal to authority would not be productive for the former situation.

Students sitting in a physics lab may also be in different frames, and thus consider different activities appropriate. For example, one student may be hoping the lab will clearly demonstrate Ohm’s law so he will remember it better, while another student may be trying to understand what voltage really is. A student looking for a demonstration would not want to spend a lot of time setting up equipment and troubleshooting – such activity is not just a waste of time, but also may leave him more confused and unsure about what to remember. A student trying to understand voltage would see setting up as an opportunity to understand how all the pieces go together, and perhaps test certain ideas and observe the results.

Dissertation Overview

In this section we state the goals and briefly summarize the main ideas of the dissertation. When helpful, we use the framework described above. To make the connections between ideas clearer, we have decided not to break this section up with headers. However, we do indicate the chapter where each idea is treated in depth.

As mentioned before, the main goal is for students to be able to use a basic understanding of measurement and uncertainty to design, run, analyze, and evaluate their own experiment and other scientific evidence. Uncertainty analysis is a tool used to convince others of a certain data interpretation. Therefore, resources associated with communication, arguing convincingly, and evaluating reasoning will be productive for students to build an understanding of uncertainty. To elicit these resources, we design the laboratory to mimic a real scientific community seeking to answer a certain question (hence the name, *scientific community laboratory*). Students must be in a frame where they are investigating a question with an unknown answer and forming a conclusion that may contradict the conclusions of other students. The physics lab is the ideal setting for teaching this, since topics in classical mechanics are related to a student’s everyday experiences in a very obvious manner, making it possible for them to invent experimental methods. Also most experiments are relatively cheap and quick to perform, making multiple trials possible. We describe the design of the scientific community labs and their purpose in Chapter 3. In a typical lab

---

[1] See Tannen (1993) for the definition of frame as it is used in the fields of anthropology, sociology, artificial intelligence, and linguistics.

class students make multiple measurements and use several different methods to answer real-world questions. While working on the scientific community labs students must defend their ideas about measurement; for example, what data to take and how to interpret it. In addition to helping build student understanding, this design allows the researcher the opportunity to see what is behind students' ideas, and from this to study how students understand measurement.

Students' ideas about the main purpose of lab and its usefulness are used to determine how effective the laboratory was at setting the frame. If students think the purpose of lab is to demonstrate physics concepts, they likely would be frustrated if different groups came to different conclusions. Such students would view that situation as a poorly designed lab that failed. They might be likely to attribute the problem to "human error" in the gathering of data instead of evaluating students' data analysis. In Chapter 4 we infer students' frames from their responses to the laboratory and then assess the curriculum's effectiveness at encouraging productive frames.

Another way to determine a students' frame is through their behavior during lab. Students must spend time sense-making and thinking about their own thinking for them to build an understanding of uncertainty. In Chapter 5 we study how students spend their time in lab by coding time spent on off-task, logistics, or sense-making behavior. This analysis tool is useful for comparing different lab designs and different student groups. Adding metacognition to the coding scheme allows one to define productive metacognition and to see whether this is fostered in the laboratory setting. We find that students in the scientific community labs spent more time sense making and used more productive metacognition compared to traditional labs.

Even if students are in a productive frame, they must be using the appropriate resources to build an understanding of the necessary concepts. To teach students, we must know what the concepts are that underlie an understanding of measurement and uncertainty. We analyze data from lab videotapes, student lab reports, and student homework and find three concepts commonly used by students when gathering and analyzing data. These concepts, two resource-like and one conception-like, are presented in Chapter 6. For the specific case of comparing two sets of data, we use a flowchart representation to look at how students' arguments become more complex and convincing. A deeper understanding of students' reasoning helps us as instructors to be able to diagnose students and respond in better ways. It also provides the researcher a framework for future studies on the understanding of data analysis.

Our final goal is for students to be able to use these (identified and unidentified) concepts to analyze data in an appropriate manner. We use a modified version of Buffler et al.'s (2001) free response survey to study quantitatively how individual students gather data and compare data sets. In previous research, this survey identified student reasoning based on 'set' or 'point' paradigms. Using the resources framework, we argue that this analysis can lead to a misdiagnosis of student reasoning. Written responses from students were compiled to develop a multiple choice survey. This survey can be used to investigate students' current understanding of data analysis or to evaluate an instructional method. Results from the survey presented in Chapter 7 show that students improve after instruction in the scientific community labs.

For students to understand measurement and uncertainty, we argue that the student laboratory must be designed to encourage students to be in a frame where they view resources used to argue and evaluate as appropriate, engage in productive behavior, use those resources to build an understanding of the underlying concepts, and use those concepts to analyze data. We make use of interviews, surveys, and video data to study each of these requirements and to evaluate the scientific community lab curriculum.

The results of this project can help instructors in their interactions with students and can aid laboratory designers. Scientific community labs have been shown to help students plan experiments, analyze data, and argue for their conclusions. The cognitive behavior analysis and measurement survey are both useful for evaluating lab designs or probing student ability. The data analysis framework may help set a foundation for future research on concepts underlying other areas of data analysis. The more we understand about students' analysis of data the better we can help students use more productive reasoning.

References

Bevington, P. R., and Robinson, D.K. (1992). *Data reduction and error analysis for the physical sciences* (2nd Ed.). New York: McGraw-Hill, Inc.

Bickley, L. S., and Hoekelman, R.A. (1999). *Bates' guide to physical examination and history taking* (Seventh Edition). Philadelphia: Lippincott.

Buffler, A., Allie, S., Lubben, F., and Campbell, B. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *Int. J. Sci. Educ., 23*(11), 1137-1156.

Duggan, S., and Gott, R. (2002). What sort of science education do we really need? *Int. J. Sci. Educ., 24*(7), 661-679.

Hammer, D., and Elby, A. (2003). Tapping epistemological resources for learning physics. *Journal of the Learning Sciences, 12*(1), 53-90.

Hammer, D. (1996). Misconceptions or P-Prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *The Journal of the Learning Sciences, 5*(2), 97-127.

Hofer, B. K. (2001). Personal epistemology research: implications for learning and instruction. *Ed. Psych. Rev., 13*(4), 353-382.

Leach, J., Millar, R., Ryder, J., Sere, M-G., Hammelev, D., Niedderer, H., and Tselfes, V. (1998). Students' images of science as they relate to labwork learning. *European project LABWORK IN SCIENCE EDUCATION* (Working Paper 4 (Targeted Socio-Economic Research Programme Project PL 95-2005).

Lubben, F., Millar, R. (1996). Children's ideas about the reliability of experimental data. *Int. J. Sci. Educ., 18*(8), 955-968.

McCloskey, M., Caramazza, A., Green, B. (1980). Curvilinear beliefs in the absence of external forces: naive beliefs about the motion of objects. *Science, 210,* 1139-1141.

McCloskey, M. (1983). Naive theories of motion. In Genter, D., Stevens, A.L., *Mental Models* (299-324). Hillsdale, NJ: Lawrence Erlbaum Associates.

Minstrell, J. (1992). Facets of students' knowledge and relevant instruction. In Duit, R., Goldberg, F., and Niedderer, H., *Research in Physics Learning: Theoretical Issues and Empirical Studies, Proceedings of an International Workshop* 110-128. IPN, Kiel Germany, 1992.

Scherr, R. E. (2002). Coordinating theoretical models of student reasoning with evidence: An example from special relativity Philadelphia, PA: (124th AAPT National Meeting)

Schommer, M. (1990). The effects of beliefs about the nature of knowledge in comprehension. *J. of Ed. Psych., 82*(3), 498-504.

Sere, M., Journeaux, R., Larcher, C. (1993). Learning the statistical analysis of measurement errors. *Int. J. Sci. Educ., 15*(4), 427-438.

Smith, J. P., diSessa, A.A., and Roschelle, J. (1993). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Tannen, D. (1993). What's in a frame? Surface evidence for underlying expectations. In Tannen, D., *Framing in Discourse* (14-56). New York: Oxford University Press.

Taylor, J. R. (1997). *An Introduction to Error Analysis* (Second Edition). Sausalito, California: University Science Books.

Tsai, C. C. (1998). An analysis of scientific epistemological beliefs and learning orientations of Taiwanese eighth graders. *Science Education, 82*(4), 473-489.

Welzel M., Haller, K., Bandiera, M., Becu-Robinauld, K., Hammelev, D., Koumaras, P., Niedderer, H., & Paulsen, A. C., S. (1998). Teachers' Objectives for Labwork. Research Tool and Cross Country Results. *European project LABWORK IN SCIENCE EDUCATION* (Working Paper 6 (Targeted Socio-Economic Research Programme Project PL 95-2005).

Chapter 2: Review of Previous Research

Introduction

In this dissertation we look at how students take, interpret, and analyze data in the physics laboratory. We do so using several different perspectives, for example, the individual student responding to a survey and a group of students socially constructing arguments. One could attempt to do this from a theoretically neutral standpoint, arguing that this project is just presenting experimental methods and results. This could be considered analogous to a student memorizing one set of rules for how to solve horizontal motion problems and a different set for vertical motion problems. This method might serve him well (especially in a traditional physics course), and he could also memorize a set of rules for projectile motion problems. However, since projectile motion is simply a combination of the two, it would be much simpler for the student if he were able to see the connections between the two motions, and use the underlying theory to combine them into projectile motion.[1] The same is true here – making use of a theoretical perspective allows one to see connections between results and also to extend results.

This chapter starts by introducing the theoretical frameworks and then using an example of student discourse to further explain them. We use these frameworks throughout this dissertation to inform the development of research questions and the interpretation of results. We next discuss previous research on laboratory design and on students' understanding of measurement and uncertainty, making use of the theory just introduced. This research forms a foundation for the scientific community lab design and for the different investigations of student understanding. Readers familiar with or not interested in the theoretical framework may wish to start with the second half of the chapter, beginning with the section titled "Research on Laboratory Design".

Theoretical Framework

Throughout this project there are situations where it is useful to view student reasoning from different theoretical stances, including misconceptions, resources, epistemology, and frames. In this section we describe these frameworks and then examine a brief example of student discourse from different perspectives, showing how each perspective highlights a particular part of the discourse and suggests a certain intervention.

Picture a one room schoolhouse in the latter part of the 19th century: students working individually and silently, writing spelling words, working sums, or reading lessons in a primer. A few students stand at the front, answering questions about the day's lesson or working out a problem on the board. Whether tacit or explicit, such a class is based on the *tabula rasa* theory of student learning. The student is a blank slate to be filled with history dates and knowledge of how to work sums and diagram sentences. The job of the teacher is to transmit knowledge to the students, and every student is expected to receive that knowledge in an identical manner.

In contrast to that is the constructivist view: a student learns by taking an experience and melding it with their previous knowledge. Therefore, the same sentence spoken during a lecture may be interpreted differently by two students with different previous experiences. Even a professor who introduces a quantum physics course by saying "this is one area where your

---

[1] An additional benefit is that the student would see physics as a unified theory instead of a collection of random facts, leading him to search for connections in other areas.

intuitions will lead you astray, so just ignore them" is acknowledging (while trying to circumvent the problem) that students enter the class with previous knowledge which affects their present learning.

Misconceptions

The misconceptions framework identifies a structure for students' existing underlying knowledge. According to this theory, students develop reasonable, well-articulated naïve theories (called misconceptions, alternative conceptions or preconceptions) from their previous experiences (Driver, 1981). Typically a researcher searching for students' misconceptions in an area has already, through informal student interactions, identified a particular difficulty students have. The researcher will first design several questions about that area in different contexts and collect student responses to these questions through surveys or interviews. She will then sort the answers into categories and search for an underlying theory that could explain the common answers in different contexts. (See, for example, McCloskey, 1983; Clement, 1982; and Finegold and Gorsky, 1991.) According to the theory, misconceptions have the following qualities:

- They are different from accepted, or expert, conceptions.
- They are strongly held and resistant to change.
- They appear consistently in 20% or more students before and after traditional instruction.
- They must be replaced, avoided, or eliminated.

Typical instruction (e.g. McDermott and Shaffer, 1992) designed to correct student misconceptions[1] follows the *elicit, confront, resolve* method. The students work through a series of questions that brings out and makes explicit their misconception. Then they are led through an actual or thought experiment or a logical argument which contradicts their misconception. Lastly they resolve the problem with a conception that is consistent. Students work through this process several times in different contexts, generalizing their conclusions. Using this method, carefully designed research-based curricula have resulted in significant conceptual gains. For example, on a question testing understanding of electromagnetic waves, Ambrose et al. (1999) found that ~ 10% of students answered correctly before and ~ 90% answered correctly after completing the tutorial.

The misconceptions framework has been criticized for failing to provide a model for conceptual change (Smith et al., 1993). According to constructivism, students use their existing knowledge to interpret new experiences, building more knowledge which they then use to interpret more new experiences. In this picture conceptual change occurs slowly and with many intermediate states. If one conceives of misconceptions as irreducible complete beliefs, to change them they must be thrown out and replaced with the new conception. If their existing knowledge is wrong, how can students use it to build correct knowledge? There is no mechanism which allows change to occur.

A second problem is that misconceptions are not always wrong and are frequently used by experts in the field (Smith et al., 1993). Misconceptions must be productive in some cases, or people would not use them. Perhaps the misconception *force causes motion* is incorrect for a bowling ball rolling across a floor. However, it is productive when used in high-friction circumstances such as driving a motor boat or the motion of bacteria in the body (when considering only the applied force, and not the more hidden drag force or when starting from rest). The discontinuity between novices and experts is not as large as the misconceptions theory implicitly asserts.

---

[1] No curriculum is designed to follow one specific theory. They are designed to work. So one cannot truly say this electric circuits tutorial was designed from a misconceptions theory, but it agrees with the main assertions of misconceptions theory.

Resources

The resources theory[1] provides an alternative structure for describing students' underlying knowledge. Unlike misconceptions, resources are more primitive and applicable to wide ranges of situations, from physical to social to economical. For example, the resource *more cause leads to more effect* (or *more is more* for short) can be applied to circuits (a higher voltage causes a greater current), parent-child relationships (more time spent together causes a deeper relationship), and grocery shopping (more money buys more food). When learning to drive, one may be taught that a stronger push on the brakes stops a car faster, but when in an uncontrolled skid, the car will stop faster if the brakes are released and pressed. In normal driving conditions, *more is more* is a productive resource for stopping the car, but when skidding, *more is more* is unproductive. When applied to a specific situation, the general resource, *more is more,* becomes what we call a *facet.*[2] For example, when applied to the normal driving context, *more is more* becomes the facet *pushing harder on the brakes causes the car to stop faster* (see Figure 2–1).

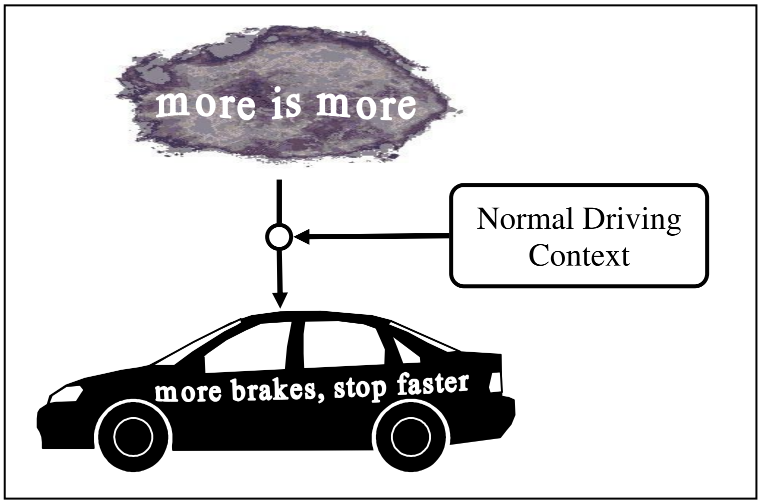


Figure 2–1

The more is more resource becomes a facet when applied to a normal driving context

The general resource *more is more* is not itself right or wrong (in contrast to a misconception), but it can be applied in productive or unproductive situations, creating right or wrong facets. Resources are generalized from a person's experience with the world and how things work. They are often directly linked to particular situations, so one situation will cue a particular resource (driving on a dry road will cue *more is more* applied to braking) but a slightly changed situation will cue a completely different resource (driving across a frozen lake, a person will apply *less is more* to braking).

A researcher wishing to identify resources may, first of all, look at the list generated by diSessa (1993). Because resources are abstracted from experience, there may exist only a limited number of them (Redish, 2003), and one may be likely to find a description of a previously discovered resource students are applying. Failing that, resources may be difficult to identify, for a student will never use the words or the abstract concept *more is more* in an explanation but just the applied facet. Resources are also impenetrable: any deeper structure or rules for use is not available to the

---

[1] See DiSessa (1993) for an in-depth discussion of phenomenological primitives, a specific kind of resource.

[2] This definition of facet is adapted from Minstrell's (1992).

user. If asked to give a reason for an answer, students may reply “that’s just the way it is”, in fact, this is a likely sign that they are applying a resource. DiSessa (1993) lists 17 principles that can be used to argue the case for the existence of a *phenomenological primitive*, one type of resource. They include the principle of unproblematic genesis: common events should exist from which the resource could be generalized, and the principle of functionality: a resource should be useful for dealing with the world. By looking at data of students generating explanations and predicting events through the filter of these principles, one can identify resources students are using.

Curricula designed based on a resources theory typically has a very different feel than that designed based on misconception theory, because of the basic principle that student’s previous knowledge is actually seen as useful for building correct ideas. One example (Hammer and Elby, 2003) comes from a lesson on Newton’s 3rd law, an area where student’s intuitions frequently seem to lead them astray. When a large truck collides with a small parked car, more than half of the students believe that the truck exerts a larger force. In any collision, 60% or more university students in introductory physics claim that the object with a larger mass or a larger velocity exerts the larger force (Bao et al, 2002). Analyzed using the resources perspective, students are applying the *more is more* resource to the collision: more mass or more velocity means an object exerts more force. They could also be applying *less is more*, by arguing that the smaller parked car feels a greater force, or more generally, “reacts more”. These resources, if applied to acceleration instead of force, result in the correct answer. Using this theoretical prediction, the curriculum is designed to elicit student’s rough intuition, apply it to acceleration instead of force, and see the implication that the forces are equal. Students are first asked to compare the forces when a truck runs into a parked car. More than 70% of students in introductory physics answer that the car feels a larger force than the truck feels. Then students are asked how much speed the car would gain during the collision if the truck loses 5 mi/hr and the truck weighs twice as much as the car. Most students answer that the car would gain 10 mi/hr. Students are then led, through Newton’s 2nd law, to see that this result would imply equal forces on the car and the truck, and that twice as much force on the car would lead to four times as much acceleration. When reflecting on their reasoning, students conclude that they need to apply their “smaller car gets more” raw intuition to acceleration, and not to force. Their intuition is productive for thinking about the situation, but it needs to be refined in a correct manner; that manner is predicted by the resources theory.

A caveat is important here: the instructional strategies given as examples for misconceptions theory and resources theory are not definitively prescribed by the theory, they are examples of how someone might teach who is using a certain theory to interpret and predict students’ behaviors. Many curricula may be theoretically neutral, or based on a combination of several theories. Curricula based on the misconceptions framework may make use of student’s previous knowledge (more of a resources trait). One example is the use of bridging analogies proposed by Clement et al., (1989). If a book is resting on top of a table or a spring, students who believe that a static object like a table cannot exert a force on the book often believe that a spring can exert a force on the book. If one bridges the two situations by placing the book on a weak, flexible board, students may see the “springiness” of the table and understand that it does exert a force. In this way, bridging analogies use students’ correct conceptions to fix their misconceptions.

Epistemology

The two previous sections used misconceptions and resources to describe students’ ideas about physics concepts. One could also consider student’s ideas about knowledge, that is, their epistemology. A person’s ideas about the source or origin of knowledge, what can be done with knowledge, and how to evaluate knowledge are all part of their epistemology.[1] Though this may seem like a fluffy philosophical issue not needed in a physics classroom, there are two reasons to discuss it here. The first reason is that students' epistemology has been shown to affect the methods they use to learn and how well they learn (Hofer, 2001). May (2002) studied introductory physics students' epistemology from their written self-reports on what they learned and how they learned it. For his sample of 30 students he found that students' ideas about what physics knowledge was and how to learn it accounted for more than 70% of the students' gains on a common physics conceptual survey.

The second reason is that measurement and uncertainty are epistemological issues. Measurement deals directly with the source of knowledge and uncertainty deals with evaluating knowledge. The main goal of the scientific community laboratory is to teach epistemology: the skills and techniques for creating, transforming, and evaluating scientific knowledge.

In the same manner as with concepts, we can describe students' epistemology as containing stable context-independent beliefs about knowledge, or as made up of general resources which may or may not be applied.[2] Research has been done based in both frameworks. Schommer (1990) proposed a system of epistemological beliefs, consisting of the following four beliefs (listed naïve → expert):

- Structure of knowledge (isolated facts → organized and complex)
- Stability of knowledge (absolute and unchanging → evolving)
- Ability to learn (innate → acquired)
- Speed of learning (quick or not-at-all → gradual)

Schommer claims that each of these beliefs may develop independently of the other, and all four need to be considered. She then used this system to develop a survey of epistemological beliefs (Schommer, 1990) containing statements such as "self-help books are usually not much help", "getting ahead takes a lot of work", and "going over and over a difficult textbook chapter usually won't help you understand it". Schommer used the survey to test the correlation of different beliefs in college students with text comprehension (Schommer, 1990) and statistical understanding (Schommer, 1992). She also studied the relationship between middle school students' epistemological beliefs and grade point averages (GPA), determining that all four beliefs had a significant positive correlation with students' GPA, even when controlling for students' IQ (Schommer, 1993).

In a similar manner to conceptual misconceptions, research based on epistemological misconceptions has been criticized for failing to provide a mechanism for change. For student's epistemological beliefs to change, students need to be building new beliefs based on their existing beliefs. But for existing beliefs to be useful, they must contain smaller elements that are productive and the misconceptions framework has not specified such an underlying structure (Hammer and Elby, 2001). Another problem comes from the context independence implied by the theory on which Schommer's survey depends. The questions in the theory are from a very general context, some unrelated to science or even the classroom, and are assumed to measure fixed epistemological beliefs. However, some studies have shown that students' epistemological beliefs change depending on the context (Stodolsky et al, 1991; Hofer and Pintrich 1997). Schommer also assumes that the determination of what is a "novice" belief and what is an "expert" belief is context independent. Elby and Hammer (2001) argue that what may be productive to believe in one context will not be productive in another. For example, the belief that scientific knowledge is certain is

---

[1] For more information on student's epistemology in introductory physics, see Hammer (1991) for an in-depth analysis of a few students using interviews, and Redish et al. (1998) for the results of a survey designed to test epistemology in a physics course.

[2] See Hammer and Elby (2001) for a discussion about epistemological concepts vs. resources.

commonly labeled naïve. If a student is studying Newton's 3rd law and believes that knowledge is tentative, then they may throw out Newton's 3rd law in cases where it seems contradictory (such as in collision described above) instead of seeking to resolve the problem. In this case, their "expert" belief has been unproductive.

The theory of epistemological resources attempts to answer these criticisms by providing a structure for describing students' ideas about knowledge. This theory claims that just as students build a general resource *more is more* from their everyday experience, they also build resources about knowledge, such as *knowledge as propagated stuff* (Hammer and Elby, 2001). This resource may be applied productively (by asking the telephone operator for someone's number) or unproductively (by copying someone's homework). Its activation depends on context, such as a class on physics or history, or even more specifically, working on physics homework or taking a physics exam. A person who wants to build an understanding of physics may be less likely to apply the resource *knowledge as propagated stuff*, and may instead apply the resource *knowledge as fabricated stuff*, as they attempt to build understanding from previous knowledge.

Frames

Which resource or misconception a person applies is partly determined by their context, of which one part is their state of mind, or *frame* (Tannen, 1993). Suppose someone needs the phone number of a friend who just moved – he could call a mutual friend or the telephone operator. Consider the resources applied during the following two conversations.

A: Hey, do you know Joanna Ramsford's number?
B: Um, yeah, wait a sec…it's 7348781638.
A: Are you sure? She just moved to Pinckney, Michigan. Isn't the area code
there 731?
B: No, it's right – I just called her last week.

C: What city and state?
A: Pinckney, Michigan.
C: What listing?
A: Joanna Ramsford.
C: The number is 7348781638.

In the first, A does not just accept the number given to him, partly because he knows Joanna just moved and because he expects the number to be consistent with an area code he remembers. This situation prompts him to check to make sure that the number is correct. In the second, these two things are still true, yet he does not check to see if the number is correct. He would not considering doing so, because it would not be productive. One does not ask the operator "Are you sure?" The situation and context surrounding the request for a phone number places the person in a certain frame which carries with it certain rules for acceptable behavior. In the *ask the operator* frame the checking resource is suppressed. In the *ask a friend* frame the checking resource is activated.

There are other differences between these two frames. When talking to the friend, A would be expected to acknowledge the help in some way, perhaps with a "thanks" and then make some sort of conversation before hanging up. If he just said "Ok, bye" that would likely be considered rude. When talking to the operator thanks is not necessary (likely a machine is reading out the number) and talking about anything else would be very strange. If the time is 4 a.m., then calling the operator would be fine, but A would need a very good excuse to call a friend. This, and many other behaviors, are all moderated by the frame.

A students' frame, as determined by the subject, the course, the activity, and other verbal and non-verbal cues, is critical for instruction. Suppose a professor wants his students to check their answer after completing a homework problem. He can tell them to do it, but when he's not there, they clearly are not checking their answers. He can make a portion of their grade on the homework dependent on whether they prove that they checked their answer. But, even so, some students still do not check. What is the problem? It may be that the students view solving homework problems as akin to asking the telephone operator for a number. If so, forcing students to check is like coming up while they are on the phone to the operator and saying "Make sure you ask them if the number is correct". A student may do so when forced, but would stop whenever the coercion was removed. For students to see checking as productive behavior, they must be in the proper frame.

The scientific community labs require very different behavior from the traditional science laboratory. Students must design their own method to take data, analyze it, and form a conclusion. Their method is judged by how well it convinced other students and the TA of their conclusion. If students are in a traditional lab frame, where they aim to prove some theory correct, expect to get the "right" answer, and talk to the TA only if something is broken, they will be less likely to use resources productive for the scientific community labs. The scientific community labs do not have a right answer. Students seeking one will compare any results obtained in the lab to the right answer. There is no need for them to be convincing, because either they have the right answer or they do not. They do not need to evaluate other group's results for being convincing. The main goal, that of being convincing, is undermined if students are in the traditional lab frame.

Transcript Example

Each of the above theories may imply different data interpretation and instructional strategies, and may predict different student reactions (Hammer and Elby, 2003). Which theory is most productive must be determined by trying it, finding the implications, and testing those implications. We demonstrate this by analyzing a transcript using these four frameworks.

The following example comes from a transcript of four students working on a scientific community lab. It has been optimized for use as an example by removing extraneous comments. In the previous lab, students investigated which properties of an object (mass, material, radius, length) affected the acceleration of the object rolling down a ramp. In this lab the students were given two cylinders indistinguishable except for an identification number. Half of the cylinders had more mass in the center, and half had more mass around the outside (see Figure 2–2), creating different moments of inertia and thus different accelerations. Each group of students must figure out if they have two of the same type or one of each type.

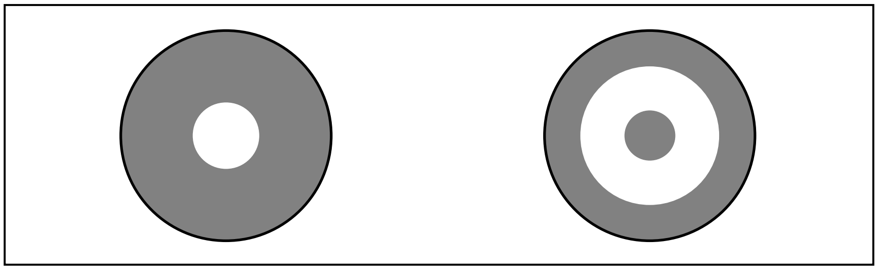


Figure 2–2

The cross-section of thick-walled and thin-walled cylinders

The students in this transcript have already spent several minutes struggling to define the question and design a method. The statements are numbered for later reference.

1 *A*: We're going to get two cans and weigh them and then see if they roll the same down the table.
2 *C*: So we're going to weigh them? Isn't that what we're trying to find out?
3 *A*: I think they have weights on them, actually.
4 *C*: They just have numbers on them.
5 *B*: We can get a scale.
6 *C*: I think the whole point is to find out if they have the same weight or not. You can't just weigh them.
7 *B*: [to TA] Do you have a scale that we can weigh the cylinders?
8 *C*: So can't we figure out from whichever one wins the race, won't that be the heavier one?
9 *B*: I think we're supposed to measure the acceleration.
10 *D*: Look at the other lab. What did we determine? We determined that [reading notes from last week's lab]
11 *C*: Acceleration equals change in velocity over change in time.
12 *B*: But then velocity, what's delta v?
13 *A*: How did we do this last week? Go back to last week's.
14 *B*: I didn't understand last week's.
15 *A*: But we just need to find it. How did we do that whole thing with the final velocity?
16 *D*: Didn't we have to multiply something by two?
17 *A*: Yeah that's what we had to do, I forget what though.

Misconceptions

A researcher or instructor using the misconceptions framework would focus on line 8, which presents student C's belief that the heavier cylinder will accelerate faster down the ramp.[1] The researcher would expect this student to also believe that a heavier object will fall straight down faster, because they would apply the same misconception to a different context. It is incorrect so the instructor would want to remove it and replace it with the correct idea, perhaps by racing two cylinders identical except for material (and thus mass) or by deriving the acceleration of two falling objects. The misconceptions framework is used in chapter 7, where we study how the individual student collects, analyzes, and compares data.

Resources

A researcher using the resources framework may also focus on line 8. However, he would interpret the student as applying the resource *more is more* to the situation, resulting in the facet *more mass means more acceleration*. Here, an instructor would want to help the student take the resource *more is more* and apply it to force, resulting in *more mass means more gravitational force*. After working through Newton's second law, the student would see the implication that two objects with the same shape but different mass have the same acceleration. The resources framework is applied in chapter 6 where we discuss students' resources for understanding measurement.

---

[1] In this section and the following sections, I do not claim that these interpretations are the only possible interpretations or that the transcript completely confirms them. This is meant as an example, so the interpretations need only be plausible.

Epistemology

A researcher considering epistemological resources might concentrate on lines 14 and 15. Student B is worried about using last week's lab, because she did not understand it. Student A brushes aside B's misgiving with "we just need to find it." One could interpret A as applying *knowledge as propagated stuff*. As long as the knowledge is in his notes, it is available and can be used. The only task, as he explains in line 17, is to find exactly where they need to multiply by two. Student B is thinking that the knowledge needs to be in her head, understood, for her to use it: *knowledge as fabricated stuff*. However, in lines 7 and 9, she seems to be applying *knowledge as propagated stuff*, by asking the TA and considering what they are "supposed" to do.

Since this dissertation is on measurement and uncertainty, by its very nature it is about epistemology: where knowledge comes from, how to manipulate it, and how certain it is. Often the idea of epistemological resources is buried when we consider more visible levels such as the cognitive behavior of a social group or an individual's use of standard deviation. In these cases it is difficult to unearth the resources a particular student is using, but also would not lead to productive information. When solving a physics problem asking about the motion of a basketball, we treat the ball as a point object instead of a collection of molecules. In a similar manner it is often useful to look at students' large-scale behavior instead of their activation of resources. We take this approach in chapter 5.

Frames

We could argue that student B is flipping between two different frames in the above transcript. While she is in a planning mode, she considers her job as one of figuring out what the TA wants her to do: whoever wrote the lab probably had a proper method in mind, and she must determine what that method is. In a traditional lab the manual would explain the method, but in these strange labs one does not get a lab manual, and thus one has to figure out the method through whatever means possible. (This may entail pestering the TA or searching the textbook for hints, methods which are often productive in a traditional lab.) However, she believes that when one is working with physics concepts, such as velocity and acceleration, one has to understand the concepts in order to apply them. She would be in this frame while working on homework or exam questions, which the velocity and acceleration issue resembles. Each particular frame comes with a set of resources she has found productive in the past and would consider applying. In this case, an instructor might wish to alter the context in lab to cue the homework frame and not the traditional lab frame. This would hopefully encourage the student to use her productive resources in planning the lab. Frame negotiation is extremely important in the design of the scientific laboratory, as we discuss in chapters 3 and 4.

Research on Laboratory Design

Having described the theoretical frameworks used in this dissertation, we move to another underlying area: the physics laboratory. All of the research in this dissertation has been done in the context of a specific laboratory design: either a traditional cookbook laboratory, or the scientific community labs. In the following section we discuss the goals of the traditional physics laboratory and give the reasons for our choice for the goals of the scientific community laboratory. We argue that the skills needed to produce, analyze, and evaluate scientific evidence are useful for students in their future, are rarely used by students before they enter the class, and need to be taught in the laboratory context.

Design Goals

Traditional goals for the introductory physics laboratory may originate from the designer's opinions, another person's opinions, or the research results of an existing lab design. Welzel et al. (1998) asked 60 teachers from 6 different European countries for their opinions about the most important goals of lab. These responses formed the basis of a standardized survey they gave to 409 teachers from different countries teaching in different subjects and school levels. This resulted in the following goal categories, of which the first three were judged to be the most important:

- for the student to link theory and practice
- for the student to learn experimental skills
- for the student to get to know the methods of scientific thinking
- for the student to foster motivation, personal development, and social competency
- for the teacher to evaluate the knowledge of the students

This is a comprehensive list; the topics are each so broad that the entire list contains any goal one might propose. The debatable point is on which of these should the laboratory focus. The scientific community labs were designed with the second and third as main goals. More specifically, the labs were designed to teach students how to produce, analyze, and evaluate scientific evidence. In the rest of this section, we build the case for this goal.

Standards for Skills Taught in the Laboratory

According to Reif and St. John (1979) there are four qualities that every skill taught in the laboratory should have. (See Figure 2–3.)

> "The laboratory should be used to teach some general intellectual skills likely to be widely useful to students in their future work. These skills should be those which practicing scientists commonly use, but which most students do not possess. Furthermore, they should be skills which can effectively be taught and practiced in a laboratory context." (p. 950)

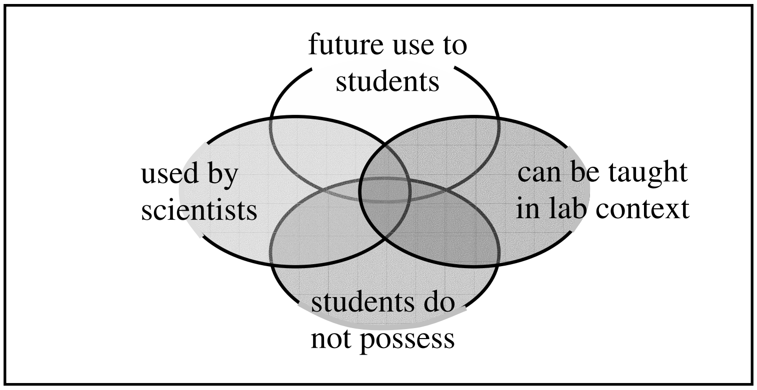


Figure 2–3

Standards for choosing laboratory skills

Most people would likely agree that these standards make sense, so now the task is to find the skills which fit them. Reif and St. John argue, and we agree, that the following skills fit these standards:

- Being able to estimate the uncertainty of quantities obtained from measurements.
- Being able to use general measuring techniques to improve reliability and precision.
- Being able to summarize the most important ideas of an experiment and elaborate as needed.

They modified the lab to give students repeated opportunity to practice these skills, with less guidance over time, and frequent assessment. Students were evaluated throughout the course with one-on-one interviews before and after lab, with half-hour exams, and with self-tests. Attitudinal surveys added to the data used to evaluate the course, collected from 250 students in the modified laboratory, and compared to students in the traditional laboratory. Results showed that students in the modified lab performed significantly better than students in the traditional course and had more favorable perceptions of the lab. For example, students were asked to summarize the main points of the lab, and then evaluated on how many essential ideas they reported (such as the purpose, theory, apparatus, main procedures, and principle results). Students in the modified course listed about 70% of the ideas deemed essential before performing the lab, and ~ 80% after completing the lab. In contrast, students in the traditional lab listed ~ 10% before, and ~ 25% after. In the use of elementary statistics, ~ 80% of the students in the modified course versus ~ 30% of students in the traditional course performed at a level deemed satisfactory by the authors. These and other data show that students did not use these skills before the course, and were able to learn them in the modified lab, fulfilling two of the four standards. More evidence that students do not use these skills is discussed in the next section on student understanding of measurement and uncertainty.

We still need to show that measurement and uncertainty skills are of future use to students and used by scientists. We argued this in chapter 1, and give more evidence here. Duggan and Gott (2002) investigated the science used by adults in the UK both in science-based employment and in everyday life. They built on previous studies which found that general science capability was at least as important as detailed conceptual knowledge (CSTI, 1993 and Coles, 1997). For science-based employment, they studied four science-based industries and a farmer, interviewing employees and senior staff and reading training documentation and protocols. For everyday life, they interviewed three parents at a local nursery about decisions regarding their child's immunization. Their results centered mainly on "concepts of evidence" which include deciding how many measurements to take over what range, how to interpret the results, and how to evaluate the experiment. They found three main areas of importance:

- People need to know and understand the principle concepts of evidence.
- People need to know how to use and apply concepts of evidence to evaluate scientific evidence.
- People need to know how to access relevant conceptual knowledge and apply such knowledge to real world issues.

These knowledge areas are beneficial to most members of society, and are used by scientists. Thus they are of extreme importance to students taking the introductory physics laboratory, many of whom[1] will be performing research on and making decisions about biological evidence directly related to our health.

Suppose one is convinced that measurement and uncertainty are important goals for the laboratory, but still wishes to hang on to the other goals mentioned previously, especially linking theory and practice (more specifically, teaching physics concepts by linking them with real world experience). These are not goals for the scientific community laboratory, but they do not conflict with the main goals. While working through measurement concepts, students will be using their

---

[1] See Chapter 3 for a detailed description of the population of students taking the algebra-based introductory physics course at the University of Maryland.

physics conceptual knowledge. Even though the laboratory is not designed to maximize students' use of physics concepts, students still gain understanding in this area.

One reason for our choice of goals is the common sense point that, considering the expense of setting up a lab, labs shouldn't try to teach topics that other components of a course can do better or with less demand on resources. For instance, two common laboratory goals, learning to follow instructions and skill in manipulation, could be taught more cheaply through a needlework class, and social skills could also be taught elsewhere (White, 1996). Knowledge such as specific laboratory techniques and how to design an investigation must be taught in the laboratory. Uncertainty analysis was developed to convince other people of one's results. Teaching uncertainty analysis outside of the lab would be like teaching swimming outside of water. Students would be unlikely to learn anything, not only because they have no chance to practice, but also because they are not in a frame which elicits the resources useful for analyzing and evaluating data and convincing others of one's analysis.[1] Some curricula have been developed to teach physics concepts in the laboratory and have been shown to do so effectively (e.g. Thornton and Sokoloff, 1990). However, these concepts can be successfully taught elsewhere. Several research-based curricula exist which aid instruction in physics concepts in the lecture[2] and in the recitation[3].

One might still ask, "Why physics?" Shouldn't these skills be taught in a course specifically designed for them, such as a laboratory methods course? Shouldn't the biologists learn this in biology? The trivial answer would be that most students do not learn it, so we need to teach it in physics. However, the subject of introductory mechanics, as compared to introductory chemistry and biology, has several traits which make it more suitable for teaching measurement and uncertainty. Measurements in introductory mechanics are relatively quick and inexpensive, making multiple measurements possible in a typical lab period. Students must be able to take multiple trials to gain information about the certainty of their data. Also, measurements in mechanics (most commonly length and time measurements) are very accessible to students, making it possible for them to determine sources of uncertainty and reason from them. But, again, one must be convinced that the skills of measurement and uncertainty are useful and important to allow the physics lab to focus on them.

Previous Research on Students' Understanding of Measurement and Uncertainty

To teach measurement and uncertainty we need to know students' initial ideas about it. Little research has been done on teaching students to interpret and analyze data. The first few papers about teaching uncertainty in the laboratory revealed the difficulties students have with the basics of taking and analyzing data. In this section, we review research on student difficulties with data analysis, and research which organized these difficulties into a framework.

Student difficulties

All of the following studies work from a misconceptions perspective: students possess reasonable, well-articulated theories about data collection, processing, and analysis. According to

---

[1] See Chapter 3 for a description of how the laboratory was designed to set up this frame and Chapter 4 for an evaluation of how well this frame was set up.

[2] For example, *Interactive Lecture Demonstrations* (Sokoloff, 2001), *Just in Time Teaching* (Novak et al., 1999) and *Peer Instruction* (Mazur 1997) are all research-based curricula for teaching physics concepts in the lecture.

[3] For examples of physics concepts instruction for the recitation, see *Tutorials in Introductory Physics* (McDermott et. al., 2002) and *Activity-Based Physics Tutorials* (Redish et al., 1997).

this perspective, the first task is to identify students' misconceptions. We describe three studies which do this.

Séré, Journeaux, and Larcher

Séré et al. (1993) studied first year university students in a physics course in France where the goals of both lecture and laboratory were to teach uncertainty analysis. Twenty students worked in pairs on an optics and an electrical resistance lab. Séré et al. observed students behavior in lab, analyzed students' lab reports and final exam results, and interviewed four students individually to see how they applied the lecture topics of measurement and uncertainty. Much of the students' behavior, according to the researchers, indicated that they did not have an underlying understanding of uncertainty analysis. For example, three out of ten groups reported a confidence interval for each of their ten trials, instead of for the whole set of data. When students used two different methods to measure the focal length of the lens, the two results disagreed for seven groups. Of these groups, four went on to report the average of the two conflicting data sets, and one group reported the integer value contained in one interval. These observations, combined with the other data, resulted in the following interpretations of behavior common to ten or more students:

- Students rarely carry out multiple measurements spontaneously, unless they doubt their first measurement.
- Students understand that more measurements lead to a better result, but they are unsure about what exactly is better.
- Students put more trust in the first value or in repeated values when evaluating a series of measurements.
- Students report only the precision of the measuring instrument when required to give a confidence interval.
- Students consider a 'bad' measurement one that has a large standard deviation, not taking systematic errors into account.

From these observations the researchers concluded that the majority of the students memorized mathematical tools and other steps for solving certain problems, instead of understanding the concepts. The researchers proposed teaching uncertainty analysis as a skill, by guiding the students through a series of situations where they can make sense of finding, using, and evaluating data. However, we make use of the framing perspective to present a different solution to the problem of students memorizing and not understanding. Students may not consider looking for an underlying theory as acceptable behavior during lab. This is similar to the previous example, where a person asks the telephone operator for a phone number and does not consider checking that the number is correct. In this example, checking is not appropriate. Likewise, in the frame of the physics lab, students may view seeking an understanding as inappropriate behavior. Students may think that their job is to follow the handout, checking off activities and answering questions. Perhaps the writer of the handout had some reason for everything, but that does not matter to the student. The guiding Séré et al. propose must take place in a frame where students will consider making sense of the situation as allowed and productive behavior.

Leach, Millar, Ryder, Séré, Hammelev, Niedderer, and Tselfes

Leach et al. (1998) built on the study of Séré et al. (1993), probing students' ideas about data analysis and the formation of theories. They surveyed a large population spread out over six European countries, including more than 400 high school students and more than 200 university students in physics, chemistry, and biology. They delivered a written survey that consisted of both open-response and closed answer choice questions, with a general context (about science) or a specific context (for example, measuring the mass of different oils). Survey results of interest showed that many students think as follows:

- It is possible to make a perfect measurement of a quantity given enough time and money (30%-60% of students).
- One should always use the arithmetic mean to obtain a final result from a set of data (80% of students).
- The average is all that matters when comparing two data sets, even if they have different confidence intervals (around 30% of students).

Their survey contained many questions that probed the same concept within different contexts. For example, one question with a general context asked students to agree or disagree that "scientists' theoretical assumptions influence their interpretation of data." Another question showed a graph of resistance vs. temperature for a superconductor. It then described how the data was interpreted by different research groups, and the students were asked to choose between research groups. One group's method was completely unrelated to theory ("Draw a line joining each of the points") and one group's method relied extensively on theory ("Consider which model could best be used to explain this data set. Once the best model has been agreed upon, a line can then be drawn through the data points.") (Leach et al., 1998, p. 84) Both questions ask about the connection between data interpretation and theory, one with a general context and one with a specific context. The different contexts of these and other survey questions allowed the researchers to determine whether any students were consistently reasoning from a central idea concerning data and analysis. They found only 14 students (about 2% of the total sample) who answered the questions consistently, showing that context plays a large part in students' reasoning.

This data seems to indicate that the different contexts are cuing different resources in the students. We can use a framing and resources perspective to suggest further research. Different question contexts could be identified which elicit more or less sophisticated responses in the students. The productive contexts could be exploited with instruction designed to make explicit the reasoning students use when answering those questions. The goal of such instruction would be for students to be able to use their productive reasoning in other contexts. The productive and unproductive question contexts could be used to generate a theory about which frames are more or less productive for students' reasoning about data analysis.

Coelho and Séré

Instead of using a paper and pencil survey, Coelho and Séré (1998) interviewed 21 French high school students performing a laboratory task. The task used an air puck with an attached sparking device which generated a series of burnt dots on a paper to determine the puck's velocity as it moved across the table. Each interview lasted about 45 minutes and involved two or three students. The interviewer ran the experiment, and then asked the students to use the dots on the paper to determine if the puck's velocity changed. While the student was working the interviewer asked questions, such as "Why do you round off figures?", "Why have different results been obtained?", "From the results of these measurements, what allows you to say that velocity has decreased?", and "Is there any point that could be considered more exact than another?" The practical context generated very similar results to the paper and pencil contexts of the previous studies. Nine of the eleven groups indicated a belief in a true value, a concept similar to what Leach et al. (1998) called a "perfect measurement". These students also tended to reject any variability in their data. They were upset when two people taking the same measurement got different results, stating, for example, "There cannot be several measurements for the same distance! You surely agree with me!" The researchers found this belief in a true value and refusal of variability to sometimes be fruitful: students who believe in a true value will search for it, trying to improve their measurement method by reducing variability. Though this idea may be considered incorrect, it can be productive in certain situations. One could also consider situations where the refusal of variability would be harmful, such as when measuring the half life of a radioactive substance. In that situation, students may spend all their time trying to eliminate the variability (which is impossible) and never be able to come to a conclusion. In certain cases, it is necessary to accept variability, and in some cases accepting variability is harmful. We discuss these different types of variation when presenting the lab goals in Chapter 3.

Frameworks for interpreting student difficulties

The previous studies were mostly designed to discover student difficulties. The following studies attempt to organize these difficulties into a framework. Such frameworks allow one to specify the goal of instruction and thus evaluate a curriculum.

Lubben and Millar

Lubben and Millar (1996) studied students aged 11 to 15 years in the UK, probing their understanding of multiple measurements with a written survey. Each survey question had a specific laboratory context (the time for sugar to dissolve or for a toy car to move 5 meters). Students were asked questions about taking and evaluating the data. Lubben and Millar compiled the results from more than 1000 students into 8 different levels. Each level is specified by three ideas: a view of the process of measuring, a way to evaluate the result, and a method for dealing with anomalous data. The lowest, level A, contains the view that a single measurement results in the true value, so evaluation of results and anomalous data are non-issues. Lubben and Millar claim that the goal of instruction is to move students to higher levels, the highest of which is level H. In this level one believes that careful measurements can approach the true value, but one cannot ever be sure of reaching the true value and so takes an average. Anomalous data can be evaluated by using the spread of the measurements and may be rejected before taking an average. (See Table 2-1.) They found level H to be common, with about 40% of students at each age level using the average to compare two sets of data.

Lubben and Millar seem to be classifying each of these levels (except the highest) as a misconception, strongly held and context independent. However, their survey asked only about two contexts (sugar dissolving and toy cars) and none of their questions had a general context. In fact, there were some cases where students used reasoning from a neighboring level if their previous reasoning no longer worked. For example, students choosing to report a repeated value (level D) may switch to the average if there is no repeated value (level E). It may be that students would not be found to be consistently in one level if the contexts differed more.

Level	View of the process of measuring	How to evaluate your result	What to do with anomalous results
A	Measure once and you get the true value.	Not an issue - a measurement is correct	Not an issue.
B	Measure once and take this as the right answer. Any result is likely to be as good as any other, so repeating is useless.	Unless something has obviously gone wrong, a measurement is correct. In familiar contexts, your result should be close to what you would expect.	Not an issue.
C	If you have adequate equipment and are careful, your measurement will be right. Take a few trial measurements to practice and then take your final measurement.	Unless something has obviously gone wrong, a measurement taken with practice will be correct. In familiar contexts, your result should be close to what you would expect.	Ignore. (Differences are due to different amounts of practice.)
D	If you have adequate equipment and are careful, your measurement will be right. Repeat trials to get the same result twice.	Getting the same value twice shows you have measured carefully enough.	Ignore.
E	Repeat a measurement and take the average. Repeating the measurement exactly will give the same result, so change conditions slightly each trial.	Variation is to be expected. Not an issue.	Variation is to be expected. Include all values in calculating an average
F	Careful measurements may be close to the true value but you can never be sure you have found it. Take an average to allow for this.	Cannot be evaluated from 'inside'. Only method is to check with an authority (i.e., teacher or textbook).	This is why we calculate an average – it takes care of the differences
G	as above	Can be evaluated from 'inside'. The spread of the measurements is an indication.	as above
H	as above	as above	Exercise judgment to reject anomalous results before averaging. The mean of some data sets may be better.

Table 2-1

Levels for students' understanding of data analysis (Lubben and Millar, 1996)

Buffler and Allie

Building on the work of Lubben and Millar, Buffler et al. (2001) developed the *Physics Measurement Questionnaire*, a survey testing students' decisions concerning data collection, data processing, and data comparison. Every question used the same experimental context, a ball rolling off an elevated ramp onto the floor. (See Figure 2–4.) When the survey was administered, the actual apparatus was used to demonstrate the experiment, so students had a clear picture of the context. Students responded to each question by choosing which statement they agreed with and then explaining their reasoning (see Figure 2–5).

An experiment is being performed by students in the Physics Laboratory.
A wooden slope is clamped near the edge of a table. A ball is released from a height h above the table as shown in the diagram. The ball leaves the slope horizontally and lands on the floor a distance d from the edge of the table. Special paper is placed on the floor on which the ball makes a small mark when it lands.

The students have been asked to investigate how the distance d on the floor changes when the height h is varied. A metre stick is used to measure d and h .

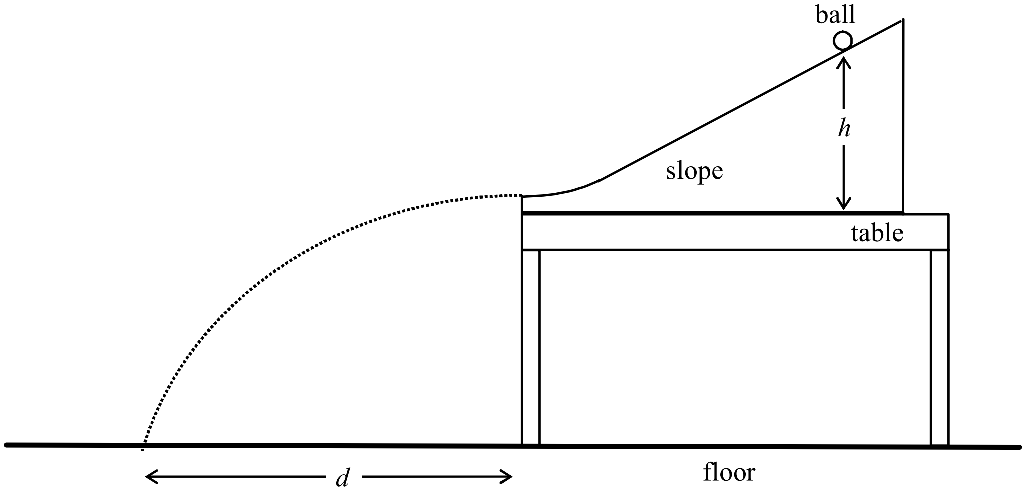


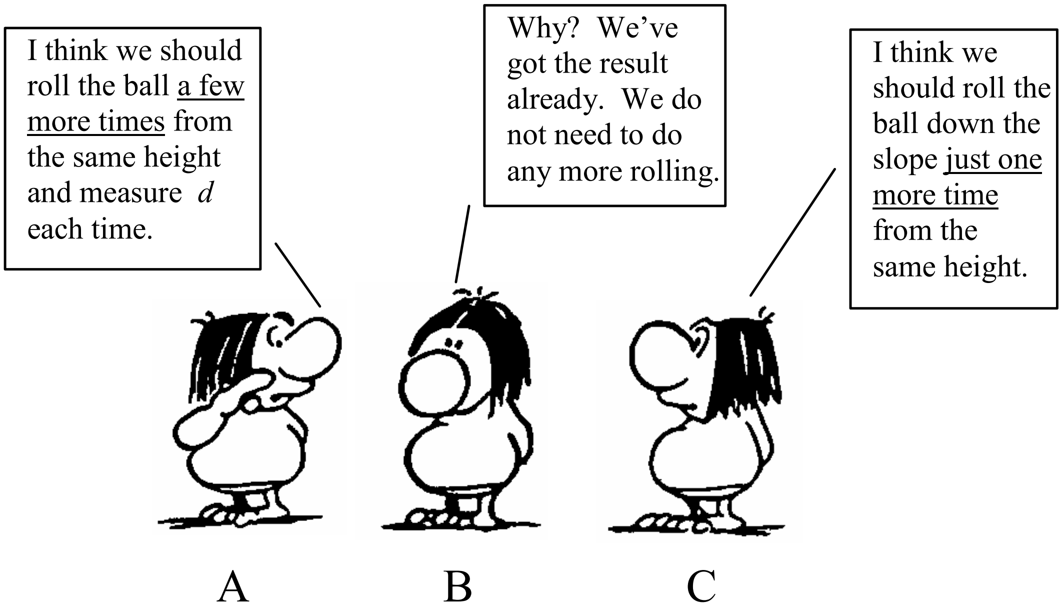
Figure 2–4

Task Context for the Physics Measurement Questionnaire
(Buffler et al., 2001, p. 1140)

Buffler et al. analyzed the responses from 70 first year students at the University of Cape Town, South Africa, before and after a physics laboratory course. Students' responses were classified as coming from two different paradigms based on whether they perceived measurement as leading to a single value (*the point paradigm*) or establishing an interval (*the set paradigm*). Reasoning from a point paradigm indicated that students thought of measurement as resulting in one true value. Students functioning in this paradigm may want to take only one measurement (similar to level A in Lubben and Millar, 1996) or they may take several measurements until they find a repeating value which they report (similar to level D). Students reasoning from the point paradigm may compare two data sets point by point to see if either of the data sets have the same values. In contrast, students reasoning from the set paradigm consider measurement as resulting in a range of values. They would want to take many measurements to calculate a better average and to determine a confidence interval. When comparing two data sets, students reasoning from a set paradigm would look at the overlap of the two ranges. Student's responses could be classified as mixed if one could not tell from their response if their reasoning was from a set paradigm or a point paradigm.

Q 1.
The students work in groups on the experiment. Their first task is to determine *d* when *h* = 400 mm. One group releases the ball down the slope at a height *h* = 400 mm and, using a metre stick, they measure *d* to be 436 mm.

The following discussion then takes place between the students.



With whom do you most closely agree?

A	B	C

Explain your choice.

Figure 2–5

Example of question from the Physics Measurement Questionnaire

Before instruction, the percentage of students reasoning from the point paradigm on the five data collection and data processing questions ranged from 54% to 77%. This decreased to 13% to 21% of the students after instruction. However, when asked to compare two data sets of five trials each with the average given, no students were coded as using set reasoning, and 98% of the students were coded as giving mixed reasoning. Most of the students answered by comparing the two averages, and they were coded as mixed because they used the concept of average but did not give any evidence that they considered all the other measurements. These students were able to use the mathematical tools of the set paradigm, but were unable to back that up with reasoning based on an understanding of the set paradigm. Buffler et al.'s instructional goal is to cause students to move toward reasoning using the set paradigm, not just using the mathematical tools. (See Figure 2–6.) In the next section we challenge this goal, arguing for a different perspective.

Students who memorize certain methods of data analysis (such as calculating an average and standard deviation) may perform actions from the set paradigm but be unable to back up those actions with set reasoning. This is similar to Séré et al.'s (1993) conclusion that students did not have an underlying theory of uncertainty. One of the data comparison questions in the Physics Measurement Questionnaire gave students the averages and standard deviations for two groups' data (and none of the individual data points). When answering this question after instruction, 57% of the students' responses were coded set, but only 23% of the students used set reasoning on the previous questions. Therefore, Buffler et al. claim that 34% of the students used the set paradigm either by rote or in an ad hoc way, because of the information given in the question. Buffler et al. state that these students seem to be harder to shift into the set paradigm than those students who reason from a point paradigm.

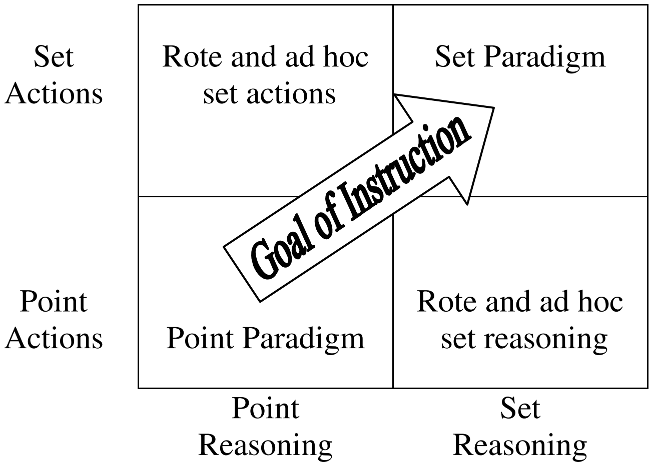


Figure 2–6

Buffler et al.'s instructional goal in relation to the set and point paradigms (Buffler et al., 2001, p. 1154)

One interpretation of this result that makes sense can be developed using a framing perspective. Students who memorize methods are not in a sense-making frame, and are not using resources productive for learning any kind of reasoning. They view their job as memorizing what the teacher wants and then using those tools when asked. Students whose responses are coded as point may be using an underlying theory to reason through their answers. To them, having an underlying theory is useful and accepted behavior. These students may be activating productive resources which will allow them to shift to set reasoning more easily.

The problem with set and point paradigms

Underlying Buffler et al.'s research is a misconceptions perspective. They wish to stop students from reasoning from the point paradigm, and teach them to reason only from the set paradigm. We would like to propose that a resources framework is more suitable. Students do not need to throw out the point paradigm and replace it with the set paradigm; instead, they need the skills to use either paradigm and to determine which paradigm to use. Different types of reasoning are productive depending on the type of question.

Suppose a coin toss determines which volleyball team serves first. There is one correct answer for the question "when you flipped that coin, which way did it land". If it landed heads, this can be considered the true value. Point reasoning is appropriate and productive in this case. It would be inappropriate for someone to use set reasoning to answer "well, it landed heads, but there's a 50/50 chance of heads or tails, so that result is not significant and I guess both teams should serve first". However, if you think you know the true value concerning which number the ball will land on in a roulette game, you will likely lose a lot of money. When considering gambling, set reasoning is productive and point reasoning is not.

It is not that point is wrong and set is right. Instead, they are reasoning tools that may be productive or unproductive to apply in different situations. The goal of instruction should not be to simply move students from point reasoning to set reasoning, but for students to be able to use either and make appropriate judgments as to which to use when. We discuss this issue further in Chapter 8, and evaluate how well the scientific community labs acheived this goal.

Summary

In this chapter we introduce the theoretical perspectives of misconceptions, resources, epistemology, and frames. A misconception is a relatively context independent belief held by a student. A resource is an idea, generalized from experience, which a student may or may not apply depending on the context. These two theories may not be contradictory: perhaps it is useful to think of students as having misconceptions in certain areas, and resources in others. Or, one could think of misconceptions as made up of resources that have a high probability of being activated in a specific context. A student could have misconceptions or resources about physics concepts or epistemology (their ideas about knowledge). The state of mind, or frame, a student is in determines the resources they are likely to activate.

These frameworks are shown useful in making sense of previous research on general laboratory goals, and on student understanding of uncertainty. We describe previous research which identified many common difficulties students have with analyzing data, including Buffler et al.'s (2001) study which classified students' reasoning in terms of set and point paradigms. Using the resources perspective we argue against their goal for students to reason only from the set paradigm; instead we argue that students should be able to use both paradigms as each is appropriate.

Finally, we conclude that it is important for students to build an underlying theory of measurement and not just memorize rules for how to treat data. Students will not attempt to build a theory unless they think such a theory exists and that it will be productive to put forth effort into making sense of the theory. The design of the scientific community labs, described in the following chapter, attempts to set up a frame where this is true (and where students recognize that it is true).

References

Ambrose, B. S., Heron, P.R.L., Vokos, S., and McDermott, L.C. (1999). Student understanding of light as an electromagnetic wave: Relating the formalism to physical phenomena. *Am. J. Phys., 67*(10), 891-898.

Bao, L., Hogg, K, and Zollman, D. (2002). Model analysis of fine structures of student models: An example with Newton's third law. *Am. J. Phys., 70*(7), 766-778.

Buffler, A., Allie, S., Lubben, F., and Campbell, B. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *Int. J. Sci. Educ., 23*(11), 1137-1156.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *Am. J. Phys., 50*(1), 66-71.

Clement, J., Brown, D.E., and Zietsman, A. (1989). Not all preconceptions are misconceptions: finding 'anchoring conceptions' for grounding instruction on students' intuitions. *Int. J. Sci. Educ., 11*(Special Issue), 554-565.

Coelho, S. M., and Sere, M-G. (1998). Pupils' reasoning and practice during hands-on activities in the measurement phase. *Research in Science & Technological Education, 16*(1), 79-97.

Coles, M. (1997). Science Education - vocational and general approaches. *School Science Review, 79*(286), 27-32.

CSTI. (Council of Science and Technology Institutes). (1993). *Mapping the Science, Technology, and Mathematics Domain*. London: The Council of Science and Technology Institutes.

diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction, 10*(2&3), 105-225.

Driver, R. (1981). Pupil's alternative frameworks in science. *Eur. J. Sci. Educ., 3*(1), 93-101.

Duggan, S., and Gott, R. (2002). What sort of science education do we really need?. *Int. J. Sci. Educ., 24*(7), 661-679.

Elby, A., and Hammer, D. (2001). On the substance of a sophisticated epistemology. *Science Education, 85*(5), 554-567.

Finegold, M., and Gorsky, P. (1991). Students' concepts of force as applied to related physical systems: A search for consistency. *Int. J. Sci. Educ., 13*(1), 97-113.

Hammer, D. (1991). *Defying Common Sense: Epistemological beliefs in an Introductory Physics Course.* Unpublished Doctoral Dissertation, University of California at Berkely.

Hammer, D., and Elby, A. (2001). On the form of a personal epistemology. In Hofer, B.K. and Pintrich, P.R., *Personal Epistemology: The psychology of beliefs about knowledge and knowing*. Mahway, NJ: Erlbaum.

Hammer, D., and Elby, A. (2003). Tapping epistemological resources for learning physics. *Journal of the Learning Sciences, 12*(1), 53-90.

Hofer, B. K., and Pintrich, P.R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140.

Hofer, B. K. (2001). Personal epistemology research: implications for learning and instruction. *Ed. Psych. Rev., 13*(4), 353-382.

Leach, J., Millar, R., Ryder, J., Sere, M-G., Hammelev, D., Niedderer, H., and Tselfes, V. (1998). Students' images of science as they relate to labwork learning. *European project LABWORK IN SCIENCE EDUCATION* (Working Paper 4 (Targeted Socio-Economic Research Programme Project PL 95-2005).

Lubben, F., Millar, R. (1996). Children's ideas about the reliability of experimental data. *Int. J. Sci. Educ., 18*(8), 955-968.

May, D. B. (2002). How Are Learning Physics And Student Beliefs About Learning Physics Connected? Measuring Epistemological Self-Reflection In An Introductory Course And Investigating Its Relationship To Conceptual Learning. Doctoral Dissertation, Ohio State University.

Mazur, E. (1997). *Peer Instruction: A user's manual*. New Jersey: Prentice Hall.

McCloskey, M. (1993). Naive theories of motion. In Genter, D., Stevens, A.L., *Mental Models* (299-324). Hillsdale, NJ: Lawrence Erlbaum Associates.

McDermott, L. C., and Shaffer, P.S. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies. *Am. J. Phys., 60*(11), 1003-1013.

McDermott, L. C., Shaffer, P.S., and the Physics Education Group. (2002). *Tutorials in Introductory Physics*. New Jersey: Prentice Hall.

Minstrell, J. (1992). Facets of students' knowledge and relevant instruction. In Duit, R., Goldberg, F., and Niedderer, H., *Research in Physics Learning: Theoretical Issues and Empirical Studies, Proceedings of an International Workshop* 110-128. IPN, Kiel Germany, 1992:.

Novak, G., Patterson, E.T., Gavrin, A.D., and Christian, W. (1999). *Just in Time Teaching: Blending active learning with web technology*. New Jersey: Prentice Hall.

Redish, E. F., Saul, J.M., and Steinberg, R.N. (1997). On the effectiveness of active-engagement microcomputer-based laboratories. *Am. J. Phys., 65*(1), 45-54.

Redish, E. F., Saul, J.M., and Steinberg, R.N. (1998). Student expectations in introductory physics. *Am. J. Phys., 66*(3), 212-224.

Redish, E. F. (2003). *Teaching Physics with the Physics Suite*. John Wiley & Sons, Inc.

Reif, F., and St.John, M. (1979). Teaching physicists' thinking skills in the laboratory. *Am. J. Phys., 47*(11), 950-957.

Schommer, M. (1990). The effects of beliefs about the nature of knowledge in comprehension. *J. of Ed. Psych., 82*(3), 498-504.

Schommer, M., Crouse, A., and Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. *Journal of Educational Psychology, 84*(4), 435-443.

Schommer, M. (1993). Epistemological Development and academic performance among secondary students. *Journal of Educational Psychology, 85*(3), 406-411.

Sere, M., Journeaux, R., Larcher, C. (1993). Learning the statistical analysis of measurement errors. *Int. J. Sci. Educ., 15*(4), 427-438.

Smith, J. P., diSessa, A.A., and Roschelle, J. (1993). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Sokoloff, D. R., and Thornton, R.K. (2001). *Interactive Lecture Demonstrations*. New York: John Wiley and Sons.

Stodolsky, S. S., Salk, S., and Glaessner, B. (1991). Student views about learning math and social studies. *American Educational Research Journal, 28*(1), 89-116.

Tannen, D. (1993). What's in a frame? Surface evidence for underlying expectations. In Tannen, D., *Framing in Discourse* (14-56). New York: Oxford University Press.

Thornton, R. K., and Sokoloff, D.R. (1990). Learning motion concepts using real-time microcomputer-based laboratory tools. *Am. J. Phys., 58*(9), 858-867.

Welzel M., Haller, K., Bandiera, M., Becu-Robinauld, K., Hammelev, D., Koumaras, P., Niedderer, H., & Paulsen, A. C., S. (1998). Teachers' Objectives for Labwork. Research Tool and Cross Country Results. *European project LABWORK IN SCIENCE EDUCATION* (Working Paper 6 (Targeted Socio-Economic Research Programme Project PL 95-2005).

White, R. T. (1996). The link between the laboratory and learning. *Int. J. Sci. Educ., 18*(7), 761-773.

Chapter 3: Context of Current Research:
the course, the students, and the laboratory curriculum

Introduction

Most of the results in this dissertation come from students working on the scientific community labs in the algebra-based physics course. To be able to interpret these results, one must have at least a brief idea of the context surrounding the research. In this chapter we describe the course, the students, and the lab curriculum. We begin by briefly describing the physics course and the modifications made for the course to use physics to teach reasoning and thinking skills. We then focus on the students served by this course: their general demographics and their expectations and opinions about physics courses. Finally we describe the design of the scientific community labs in enough detail for someone who may be considering implementing the curriculum. The goals of the lab, including often ignored measurement concepts, are explained along with particular aspects of the lab design and how they meet the goals. We also describe how the design attempts to create an environment which encourages students to be in a productive frame. (We evaluate this attempt in the next chapter.) The labs are led by graduate teaching assistants needing training on interacting with small groups and the whole class, so we end with a description of the TA training.

Algebra-based Introductory Physics Course

Course Logistics

The weekly laboratory is a required component of the algebra-based introductory physics sequence at the University of Maryland. The course consists of the following components (per week):

- Two hours of laboratory (20-24 students, one TA).
- One hour of discussion section (20-24 students, one TA).
- Three hours of lecture (120-160 students, one professor).

The course meets for fourteen weeks and typically covers the following concepts: kinematics, Newton's laws, momentum, work and energy, circular motion, torque, gravitation, heat and temperature, and oscillations.

Meta-learning course general modifications

The scientific community labs comprise one environment of many developed to help foster science learning skills as part of the project *Learning how to learn science: Physics for bioscience majors*[1]. The skills of learning science (which we will call *meta-learning skills*) include shopping for useful ideas, identifying real life experiences which embody scientific principles, tracing out the implications of an idea, reconciling one's raw intuition with scientific theory, and seeking consistency between two ideas. In addition to the labs, we also modified the lecture, the homework assignments and the discussion section. In this section we briefly discuss those changes.

---

[1] Supported in part by NSF grant REC-008 7519

Lecture

Lecture was changed to make it more interactive and engaging and also to teach meta-learning skills alongside physics concepts. One way to increase student involvement and learning in the large lecture setting is through *interactive lecture demonstrations* (ILDs) developed by Sokoloff and Thornton (1997 and 2001). These are a series of demonstrations illustrating fundamental physics principles, often using computers to take and display data in real time. The instructor explains the demonstration without showing the results and asks students to predict the result individually and then discuss their prediction with their neighbor. Then the instructor asks for students' predictions and their reasons, runs the experiment, and discusses the results. Students are given a worksheet to fill out with their individual prediction and with the final result. In the meta-learning class, ILDs were modified to teach thinking skills as well as physics concepts. The purpose of a typical ILD discussion is to identify the correct answer and explain why it is correct. The modified ILD discussion sought to explore the whole space of possible answers, to propose believable wrong answers, and to explain why someone might believe them and why they are wrong. This discussion gives students the opportunity to use skills such as seeking consistency between ideas.

We assigned an infra-red remote answering device[1] to each student. This requires every student to respond to questions instead of just those students sitting in the front rows of the lecture hall, and clearly communicates to the students that we want them participating and talking in class. TAs who taught lab for several years noticed an increase in students' willingness to talk in front of the class in lab after the devices were implemented in lecture. The modified ILD's and the infra-red remotes may have changed the students' frame in class. Talking in front of everyone was no longer delegated to only the bold and brilliant: students could propose right or wrong answers without taking credit for them, and the whole-class poling allowed a student to see that she was not the only one with a certain idea.

Homework assignments

Homework assignments were changed to remove any plug-and-chug type problems and add in new types of problems: representation-translation questions (ask students to interpret various graphs, tables, pictures, etc into other forms), context-based reasoning problems (questions based in a real-world context), estimation problems (use everyday knowledge and aim for an answer with a 90% certainty), qualitative questions, and essay questions[2]. Essay questions and context-based reasoning problems often asked students to reflect on their learning process or to explicitly use one of the meta-learning skills that the class was attempting to help students develop.

Discussion Section

During the traditional discussion section a TA stands at the board solving homework problems while students sit at their desks taking notes and asking questions. This was replaced with students working in groups of four while a TA roams around giving guidance as needed. The students worked on a *University of Washington (UW) style tutorial*: a worksheet whose questions led them through difficult physics concepts in a guided inquiry process (McDermott and Shaffer, 2002). Some UW tutorials were modified and new tutorials written to teach both a physics concept and a certain meta-learning skill useful for learning that concept.

---

[1] For more information on remote answering devices, including how to implement them and their results, see Burnstein and Lederman (2001), Poulis et al. (1998), and Shapiro, (1997).

[2] See Chapter 4 in Redish (2003) for more information on these types of homework problems along with examples.

Laboratory

The traditional introductory laboratory at the University of Maryland requires students to work in pairs following the steps detailed in a laboratory manual. We replaced this "cookbook" laboratory with a more open-ended lab where students are given only a paragraph describing the research question and must work in groups of four designing and carrying out an experiment to answer that question. At the end of lab each group presents their method, data, and analysis to the rest of the class, trying to convince the other students of their conclusion. The design of the lab attempts to model a scientific research community, hence the name, *scientific community laboratory.*

Because all of these changes were made simultaneously, it is very difficult to disentangle any one cause for an effect. However, precisely because we are teaching meta-learning, instruction had to permeate the whole course. Encouraging students to look for consistency in lecture and then not rewarding it in homework would be an exercise in futility. These changes were implemented in the Fall 2001 and Fall 2002 semesters with 120 and 160 students, respectively.

Student Population

The algebra-based introductory physics course is a service course designed for specific students. These students come into the course with a certain physics background, math background, attitude, and other characteristics different from the population of other physics courses. Information on the student population will be used to explain certain aspects of the course design, and to interpret data taken on the students.

Student demographics

At the University of Maryland physics is a requirement for the biological sciences major so a large part of our students (50-60%) major in this area. (See Table 3-1 for area specializations.) Students in any major may take this course to fulfill their University science laboratory requirement, so we've had Spanish, Math, and even a Dance major take the course. Though this course is often referred to as "pre-med physics" only 30-40% of the students plan to apply to medical school. Biology students usually wait to take it until their junior or senior year because they do not need this course as a prerequisite for any of the required courses in their major. Algebra is the only prerequisite for this course; however, more than 95% of the students have successfully completed two semesters of calculus. Even with this coursework they do not feel confident in their math skills (as shown by anecdotal evidence and as self-reported by the students). About 60% of the students are female. This gender split is rare for a physics course and coupled with the general low confidence level results in a very different class culture than the typical over-confident male-dominated physics population found in physics major or engineering physics courses.

Class Standing	Fall 2002		Major Specialization	Fall 2002
Freshman	7%		Physiology and Neurobiology	28%
Sophomore	20%		General Biology	17%
Junior	58%		Cell Biology and Genetics	15%
Senior	14%			

Table 3-1

The year and common majors for students taking algebra-based introductory physics

Students' responses to introductory physics

The algebra-based physics student population has certain expectations for any course, and specifically for a physics course. In this section we describe student responses to the traditional physics course and the meta-learning course. Students' responses will give us information about the frame they are in when attending class, which will be helpful for interpreting their responses to the laboratory curriculum[1] and other research results. For this kind of initial information gathering, the semi-structured interview (described in the next section) is most useful, as it allows the researcher to tailor questions according to the subject's responses and investigate any particular topic more deeply.

The interview process

The semi-structured interview is a research tool frequently used to deeply probe the ideas of a few students. Typically one interviewer sits in a room with one human subject and a tape recorder and begins with a predetermined list of questions, called the protocol. The interviewer must first make the subject feel comfortable and encourage them to talk freely. For this to be possible, the interviewer must be someone who does not have any affect on the subject's grade, and the interview itself should have no relation to the subject's grade. After the subject has offered a response to the first question, the interviewer must choose which piece of that response to pursue further by asking a clarification or follow-up question. Obviously, the interviewer's research interests and his interpretation of the student's responses will affect the results. One could argue, however, that this is true for any kind of research[2]. After the interview the recording is transcribed and analyzed using same-gender code names.

The general purpose of these interviews was to investigate what students thought about all the aspects of the physics course. I identified myself as being involved in re-designing the physics class and asked students to volunteer to talk about their experiences. Volunteers were usually high performing students willing to help me or extremely frustrated students hoping to prevent future students from this torture. During the interview students felt willing to vent their frustrations freely and offer suggestions. We report the results from two sets of interviews here. The first set consisted of 10 students interviewed after the first exam in a traditional first semester mechanics course. In the second set, 13 students from the new meta-learning course were interviewed at the end of the second semester. Eight of these students had completed both semesters (mechanics, and electricity and magnetism) in the meta-learning section and five attended a traditional mechanics semester and then switched into the meta-learning course for the second semester.

Each set of interviews had its own protocol, and the full versions are included in Appendix D. Students were asked what they found useful about each component of the course (lecture, lab, discussion, homework, and exams), how they prepared for each component, what they found different, and what they would change. Their responses were transcribed and analyzed. Relevant themes found in the interviews are described below.

---

[1] Student's responses to the traditional physics lab and to the scientific community lab are analyzed in Chapter 4.

[2] See Ericsson and Simon (1996) for an in-depth discussion of verbal protocol analysis, including different methods of collecting and analyzing verbal data.

Results

It's strange to be asked to think

In the meta-learning interviews twelve out of the thirteen students talked about thinking (in contrast to memorizing) at some point during the course of the interview, but only one of the ten students from the traditional class mentioned it. Students who mentioned thinking defined it as the following things (same gender code names in parenthesis, see Appendix D for more information):

- "reasoning and logic" (Brian)

- "Rather than just pulling out numbers and plugging them in and not knowing what you're plugging them into or why, it's understanding what you're doing. …It's so conceptual but at the same time it's like application." (Sharon)

- "start from the beginning and use what you know, sort of put together a picture… You know, you got to use your brain, it's not just memorizing, you really have to think about things and how they work, yeah, so, I guess reasoning skills" (Henry)

- "break down into really simplistic ideas just to figure out what's going on and use that to predict what's going to happen in simple relationships, like our mechanics and our cars crashing and what not, and then you can build upon it from there" (Jacob)

Ten of the thirteen students who mentioned thinking agreed that thinking is a good thing. One student said "it's good to think about the things too…not have them just presented to you" (Henry). Students believe that thinking is good in general, but they do not always do it. Three out of the thirteen meta-learning students commented that thinking is not a normal activity in other classes.

> "You have to be ready to think, be ready to apply, everything. Whereas in other classes it's kind of like okay that's the answer. But [in this class] you have to think and apply." (Sharon)

Students believe that thinking is good, but they will not think automatically; they will only think if they are convinced it will be a productive activity. The external environment of the class must encourage students to be in a frame where they will be open to thinking. One might believe that the only way to reward thinking would be through the course's grading policy. However, students have other motivations – not all students take this course with a good grade as their only goal. Two of the 13 meta-learning students mentioned that thinking is not necessarily positively correlated with grades, and one of them specifically switched into the meta-learning class to learn more.

> "[First semester] I got an A on that class but I didn't feel I was learning anything…And that's what I wanted to get, that way of thinking about things without just using straight equations and memorizing them….I'm learning more this semester than I did last semester but I don't think grade wise I'm doing better." (Joshua)

What characteristics of a course create a frame that encourages students to think? Students themselves commented that they were willing to think if they found physics useful. Two of the 13 meta-learning students did not find physics useful, and were unhappy with the thinking they were expected to do.

> "Because it's required for our major but it's one of those classes we're only going to use a very little bit of it, so I might have gone with a professor who required less, just because I can't lie." (Brian)

Students who think physics is a useless requirement will not be willing to put forth effort unless it is clearly correlated with their grade. Fortunately, this opinion showed up rarely in the interviews. Ten of the 13 meta-learning students found physics useful and mentioned that they were willing to expend that effort.

> "I put a lot of effort into this class, more than I've ever put into any other college class. But I learned a lot more; I took more from it than I have any other college class too, so it was worth it." (Sharon)

Now the questions remain of which parts of the course do students find useful and how to convince students that physics is useful.

Physics may/may not be useful in the future

In both of the sets of interviews, we asked students if they expected that they would use anything from this course two years after they graduate. Six of the ten traditional students answered no, it is just a hurdle over which they must jump. Seven of the 13 meta-learning students mentioned that the physics concepts would be useful in biology, listing radioactivity, torque, and x-rays as examples. Students also mentioned general skills emphasized in the meta-learning course that would be useful, such as problem solving and communication skills.

Thinking that physics is useful is not enough, unfortunately. One of the traditional students saw the potential use of the physics concepts, but did not feel she understood enough to apply them.

> "like biomechanics, with the torque around your wrists and the pressure that you put on a bone that causes it to break…. I could see how this could be relevant if I know it, but since I don't, I'm just praying that it's not going to be relevant." (Liz)

Seeing the usefulness of physics may be necessary, but it is not sufficient. The student needs to be able to use resources productive for learning physics. The meta-learning course was an attempt to design the whole course (lecture, discussion, and lab) to teach and encourage the use of these resources. In the next section we discuss how the design of the scientific community labs encourages the use of resources productive for learning measurement and uncertainty.

Description of the Scientific Community Labs

Having discussed the general course design and the intended lab participants, we now focus on the scientific community labs. The scientific community labs are a series of research questions designed to help students build an understanding of measurement and uncertainty. Each research question elicits new measurement ideas and revisits ideas discussed previously. We begin by explaining the goals of the lab and how the lab is designed to meet those goals.

Goals

The main goal of the scientific community labs, argued in chapter 2, is to teach students how to produce, analyze, and evaluate scientific evidence. Students must understand the concepts underlying uncertainty in an experiment (called *measurement concepts*) and be able to use that knowledge to design an experiment and interpret their data. Measurement concepts are analogous to the physics concepts of normal force or distinguishing speed from velocity: as those concepts are useful and often necessary for building an understanding of force and motion, so measurement concepts are useful and often necessary for building an understanding of measurement. The main measurement concepts are described below.

New Measurement Concepts

There are two measurement ideas that deserve special recognition because they are so frequently overlooked. This can lead to communication problems between students and teachers, in the form of teachers misdiagnosing students and students misunderstanding questions from teachers. In this section, we explain these two concepts and why they are important.

Internal vs. External Variation

When students notice a range in their data they are quick to attribute it to "human error", deciding that something must have gone wrong. This can cause problems during lab. During a radioactivity lab one semester, a group of students was measuring the counts from a short-lived source in order to determine its half-life. They were distressed by the variation in counts, especially when the number of counts per minute increased instead of decreasing. They interpreted this as error, blaming it upon the person timing, "breezes in the room", or the Geiger counter, and spent an hour of lab trying unsuccessfully to fix the problem. This and many other ranges are often not caused by error or by something wrong in the design or implementation of the experiment. They come from something internal to the system being measured. We call variation such as the range in decay counts *internal variation*, as opposed to variation such as the range in times read off a stopwatch, which we call *external variation*.[1]

We explain these ideas further with the help of students' answers to a laboratory quiz, given in the Fall 2002 semester to students in the scientific community labs. In this question, students were asked to compare two sets of data to see which one was larger. (See Figure 3–1.)

Which battery lasts longer, Energizer or Duracell? A student performs an experiment measuring the number of hours two AA batteries from each brand will run a tape player. Her data is below.

	Trial 1:	Trial 2:	Trial 3:	Trial 4:	Trial 5:	Average:
Duracell (hours)	11.4	12.2	7.8	5.3	10.3	9.4
Energizer (hours)	11.6	7.0	10.6	11.9	9.0	10.0

Figure 3–1

The battery lab quiz question

About half the students answered this question by comparing the two averages, for example, "The Energizer battery lasts longer because, from the 5 trials, it had a longer average of hours

---

[1] We have been unable to find mention of distinguishing internal/external variation in any literature. The terms used in the field of psychology, external validity and internal validity (Campbell and Stanley, 1963), are unrelated.

lasted." A more sophisticated answer might take into account the spread of data, and recognize that the data points overlap almost completely so the difference in averages is not significant. (See Figure 3–2.) One example of such reasoning was "The information shows there is significant overlap. So we cannot really tell which battery lasts longer. They last the same approximately."

If we look more closely at the last two sentences in that answer, they could be interpreted as being contradictory statements.[1] The first, "we cannot really tell which battery lasts longer", claims that the large range is caused by the measurement process. If that range were smaller, perhaps the two data sets would not overlap as much and one could see a difference. Perhaps when the student tested the batteries she was not careful to keep everything constant between trials – perhaps she played different music at differing volume. It may be that Energizer lasts longer, but one cannot use the data resulting from this method to make that claim. Because the range is something caused by the researcher's experimental design, it is external variability.

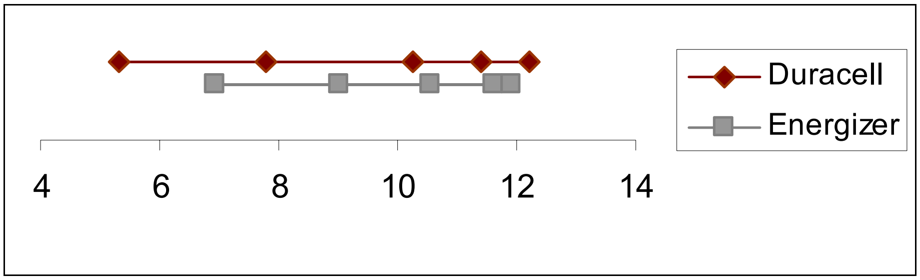


Figure 3–2

A representation of the data in the battery lab quiz question. The large overlap indicates that the difference between the averages is not significant.

The second statement, "They last the same approximately", claims that the range is caused by the batteries themselves. Some batteries the student tested just happened to last longer than others. This may be because the manufacturing process is variable, or perhaps the batteries were stored in different conditions, or are different ages. The range is inside the system which is being tested, so it is internal variability. In contrast to external, internal variability cannot be judged harmful or helpful. If one were in charge of quality control at the battery manufacturing plant, one would specifically wish to measure the battery's internal variability while minimizing external variability as much as possible.

Predictive vs. descriptive questions

Another often overlooked measurement concept focuses on the difference between a question which asks for a prediction of what is likely to happen and a question which asks for a description of what already took place. This can be illustrated by the following two student answers:

> Student A: "I would say Energizer because 3 out 5 trials [sic], it lasted for longer hours than Duracell. Also, the average shows it lasting longer by .6 hrs.
>
> Student B: "Although on average for the above trials, Energizer appears to last longer, it cannot be concluded that this is the case for most or all situations because in 2/5 trials Duracell was longer and 3/5 trials Energizer was longer."

---

[1] This student's answer is being used as an example to illustrate a certain measurement concept. This may not necessarily be the student's thinking. But a teacher or another student could possibly interpret the student's answer in this way.

These two students use the same reasons (Energizer winning 3 out of 5, and having a longer average) to come up with very different conclusions. What did they do differently? Student B's statement "it cannot be concluded that this is the case for most of all situations" is answering a question asking for a prediction of which battery will last longer. He claims that we cannot predict this; it could go either way. Student A is answering a question asking for a description of what happened when the student tested the batteries. Like a chess match or soccer tournament where best out of five wins, it is true that for those five trials Energizer won. The same question, "Which battery lasts longer?" can be interpreted as a *predictive* or a *descriptive question*.

Confusion between a predictive or descriptive question can cause problems in lab. Suppose a teacher asks a question, "Does a cylinder with a larger radius roll down the ramp faster?" A student may interpret that question as a descriptive question, and answer "Yes, when we raced them, the larger cylinder won three times, the smaller cylinder only won twice." The teacher might respond "But is that significant?" This question means nothing to the student – why does significance matter when that is what happened? The teacher has misdiagnosed the student and the intervention will likely lead nowhere. Instead, the question "if you had an option to bet your life savings on the race, would you?" might switch the student to a predictive mode.

Applicability to Biology

These measurement concepts are not only useful in the physics laboratory; they are also needed in many areas of biology research. Suppose a scientist is testing water quality at a swimming beach. If the water is safe, it should have a low level of fecal coliform bacteria. He takes a 100 ml water sample and filters it, so any microbes in the water are caught in the filter. He then places the filter in a Petri dish with broth, and incubates it. The microbes grow into colonies, which can be identified and counted. Commonly used standards state that 200 colonies per 100 ml is the safe limit for swimming. The scientist runs two tests and counts 163 and 234 colonies – what should he do? Perhaps his testing method was not standardized. Were both Petri dishes incubated for the same amount of time at the same temperature? Was the equipment completely sterile? These would be sources of external variability that would negatively affect his results. Or it could be that the two water samples he took had different amounts of bacteria in them. He could take several samples from the exact same location and test them along with samples from other places at the beach. Whether it's internal or external variation will affect how the scientist interprets his data.

Now suppose another scientist is using the same test to compare the levels of two different beaches. If she is describing the test results for perspective swimmers (answering a descriptive question), she may end up recommending a certain beach. If she is predicting the bacteria count of the water (answering a predictive question), she may end up with a different conclusion. This is just one example, but there are many other cases in biology and other sciences where the concepts of internal and external variation and descriptive or predictive questions are important.

Students may be answering a predictive question and still not look at the range overlap.

> "Most likely the Energizer will. Not only is the average time higher for the Energizer, but the range of times is smaller (Duracell: 5.3 – 12.2: 6.9 hour range, Energizer: 7.0-11.9:4.9 hour range) so it's less likely to die early."

It seems to be clear that this student is answering a predictive question, and she even takes the ranges into account. This measurement concept may be necessary, but it is not sufficient. There are many other ideas we want students to understand.

Measurement and Uncertainty Goals

In addition to the new measurement concepts described above, students need other knowledge to build a full understanding of measurement and uncertainty. They must understand uncertainty in an experiment, and be able to use that knowledge to design an experiment, interpret their data, and critique the method and analysis of their own experiment and other groups' experiments.

Understand uncertainty in an experiment

When students do not get the same result for every measurement trial, the first thing they should do is look for a mechanism causing this range. This is often a knee-jerk response from students, the random list of "error" at the end of a lab report, and just by itself is rather useless. From here students need to separate out which is internal variation and which is external, and be able to predict what effect the mechanism will have (random or systematic, large or small). They will be unable to remove every source of external variation, but should be able to identify the largest source. Using all these skills, they should be able to predict the certainty of their final result, as long as they have experience with their data taking method. (For example, predict how well they can measure g by timing the period of a pendulum.)

Experimental Design

Students need to be able to design their experiment to minimize the sources of outside variability. Often this entails an artist-like manipulation of effects. For instance, it is better to time a cylinder rolling down a longer shallower ramp because it is easier to measure the longer time (especially when using a stopwatch)? Or will the friction and starting effects outbalance this benefit? If one times 10 periods of a pendulum and then divides by ten (what we call stacking), perhaps the decrease in amplitude of the swings will affect the result.

Interpret Data

To make any progress in understanding uncertainty, students must first see that it is useful to take multiple measurements. Then they need to have a rough idea that multiple measurements can give useful information about the certainty of one's data. Most students enter lab very comfortable with the idea of average, and are almost too willing to use it. They need to understand how average is useful, what range means, and how to combine an average and a range in order to make a useful report of data. They also should be able to propagate this range through any calculations they make, so that any result they report comes with a range. When comparing two sets of data, students should start out by looking just at the range overlap. When data allows, they should be able to make a histogram and look at how the two peaks of data overlap. This can lead to an understanding of what to do with the data on the tails of the distribution (the low probability data).

Critique Method and Analysis

One of the useful skills mentioned by Duggan and Gott (2002) is for people to know how to use and apply concepts of evidence to evaluate scientific evidence. Our students should be able to use the skills mentioned previously in critiquing not only their own experiments but also other scientific evidence. Is hormone replacement therapy a good idea? Should I create a 'safe room' in my house with duct tape and plastic? Is an SUV a safer vehicle? For this to happen, students need to see these skills as useful not only for physics data in the lab, but for any kind of data anywhere.

How goals fit with the laboratory design

The different aspects of the scientific community labs have been designed with these goals in mind. Because these labs are so different from traditional labs, we need to be extremely careful to clearly communicate what we expect from students and why. Students will first enter lab in a traditional lab frame, and will interpret everything in lab through that lens. Hopefully we will be able to put students in a more productive frame so they will use resources that are helpful for building understanding. The classroom as scientific community is a major part of that frame, because uncertainty analysis is useful primarily as a way to convince other scientists in your community of your results. A student learning the techniques of uncertainty analysis without using them to convince her colleagues is like learning to sail without ever stepping in a sailboat. Put in the best light, it is really boring. Put in the worst light, students will be in the frame "guess what the teacher wants", will learn rules for mimicking behavior and understand nothing. In the following sections we describe certain aspects of the lab design and how they produce an environment which encourages students to be in a productive frame.

Introductory discussion

Most labs begin with a short class discussion meant to elicit the main measurement concept for that lab. For example, the second lab begins with the TA asking about whether a bus that arrives at 9:03 is the bus actually scheduled for 9:18 (Figure 3–3).

1. A student takes the 9:18 bus every morning. One morning a bus arrives at 9:03. The sign on the bus is broken, so she doesn't know which bus it is. Is this bus her bus, or another bus?
2. According to her watch, in the last two weeks the (supposedly) 9:18 bus has arrived at 9:12, 9:21, 9:17, 9:08, 9:07, 9:28, 9:19, 9:16, 9:23, and 9:10. A bus arrives at 9:03. Is this her bus, or another bus?

Figure 3–3

Questions for the introductory discussion in Lab 2,
meant to show students the purpose of multiple measurements.

When asked the first question, students responded with many and various answers, such as "we can't tell", "buses are always late, never early, so it couldn't be her bus", and "she should just ask the driver". After given the additional data, a few students answered "she needs a new watch", and "we still can't tell". But most students used that information as an indication of the reliability of the bus. Some concluded that it could be her bus because the times are so spread out the bus could come any time. Some concluded that it was not her bus because 9:03 was a lot earlier than any of the previously recorded times. Either way students were encouraged to realize that multiple measurements are useful, which was the main point of the discussion. They also were primed for the more subtle point that common sense reasoning is encouraged and productive in this lab.

Lab handout

The lab handout begins with a real life context which introduces the research question. Similarly to the previous example, this is meant to cue resources students already use outside of the physics classroom. It is also designed to show that physics and uncertainty analysis are useful in everyday situations. For example, in lab 4 students investigate which properties (mass, diameter, length, material, etc) of a rolling object affect its acceleration. The context for this question is a pinewood derby (Figure 3–4).

Lab 4: The wheels on the car

A child you know is entering the pinewood derby in her club. She has to design and build a small wooden car for racing down a ramp, and needs help deciding what kind of wheels to order. They come with a lot of options for wheel design – you get to choose the material, shape, and size of the wheel rim. You wonder if any of these things will affect how quickly the wheels accelerate down the ramp, and thus if the car will win the race.

Figure 3–4

The real-world context for Lab 4.

Traditional labs often use equipment carefully designed for one particular lab, which sends the message that physics can only be seen by specialized scientists with specialized equipment. The real-life setting and everyday equipment (mainly rulers and stopwatches) attempt to communicate the opposite. They also convey that our goal is a strong, solid argument other students could understand, like one used in a debate with a colleague in the hallway.

The research question is an open-ended one: the answer is usually unknown, and several different experimental methods could be used. Frequently, even in cases where theory gives an answer, the experimental result is contradictory because of effects disregarded by the theory.[1] Students must develop their own method, which achieves the basic goal that students understand what they are doing and why. Without this understanding, they will not be able to reason about the certainty of their results. This also helps to set the frame that one is expected to think in these labs, instead of just following written steps.

The research question of each lab is designed to introduce one or two new measurement concepts and revisit previous concepts. (See Figure 3–5) The concepts are legitimately important for answering the research question, and frequently the students themselves first mention them as a solution to a problem they encountered. When students do not, the TA brings up the idea by using an example from one group's data or method. This allows the students to see the purpose of each measurement concept and have ownership of the ideas.

Weekly log

After lab each group of students hands in a report, which we call a weekly log. The name is chosen because it is not meant to be a report of finished work. It is a log of research activity, written so as to be useful to another student wishing to do the same experiment. As such, students are supposed to report the route they took toward making sense, dead-ends included. Writing down their path helps students be explicit about their thinking and evaluate their problem-solving method to hopefully improve it.

The log is graded out of 10 points[2], and three points are given for journaling their process of coming to a conclusion. Another three points are given for persuasiveness – how well the data supports the conclusion. As mentioned above, this is where uncertainty analysis becomes important, when one is attempting to convince a colleague to make a certain conclusion from the data. However, since these labs are similar to actual research, groups often fail to get any usable data, or are unhappy with their data quality. A group without any conclusion can still receive full credit for persuasiveness if their evaluation section makes up the difference. Students are not

---

[1] For example, in scientific community lab 2, often students find that the period of a pendulum does depend on mass, especially for very small masses. This is because they use string which has a mass comparable to the hanging mass, so as the hanging mass changes, the center of mass of the pendulum changes location.

[2] The grading scheme is described in the Lab guide handout in Appendix A.

graded for getting the correct answer[1] – if this were true, they would not have any reason to listen to other group's reports and evaluate their own lab design. They also would not be free to experiment and try out new methods. The evaluation section gives each group the chance to use information from other groups, and is also worth three points. It is where the community becomes useful and necessary: students are expected to incorporate ideas and suggestions from other groups in their evaluation. If they decide that their experiment does not need changing, they must explain why other group's methods or suggestions will not work. If their results contradicted other groups' results, they should explain why and suggest a resolution.

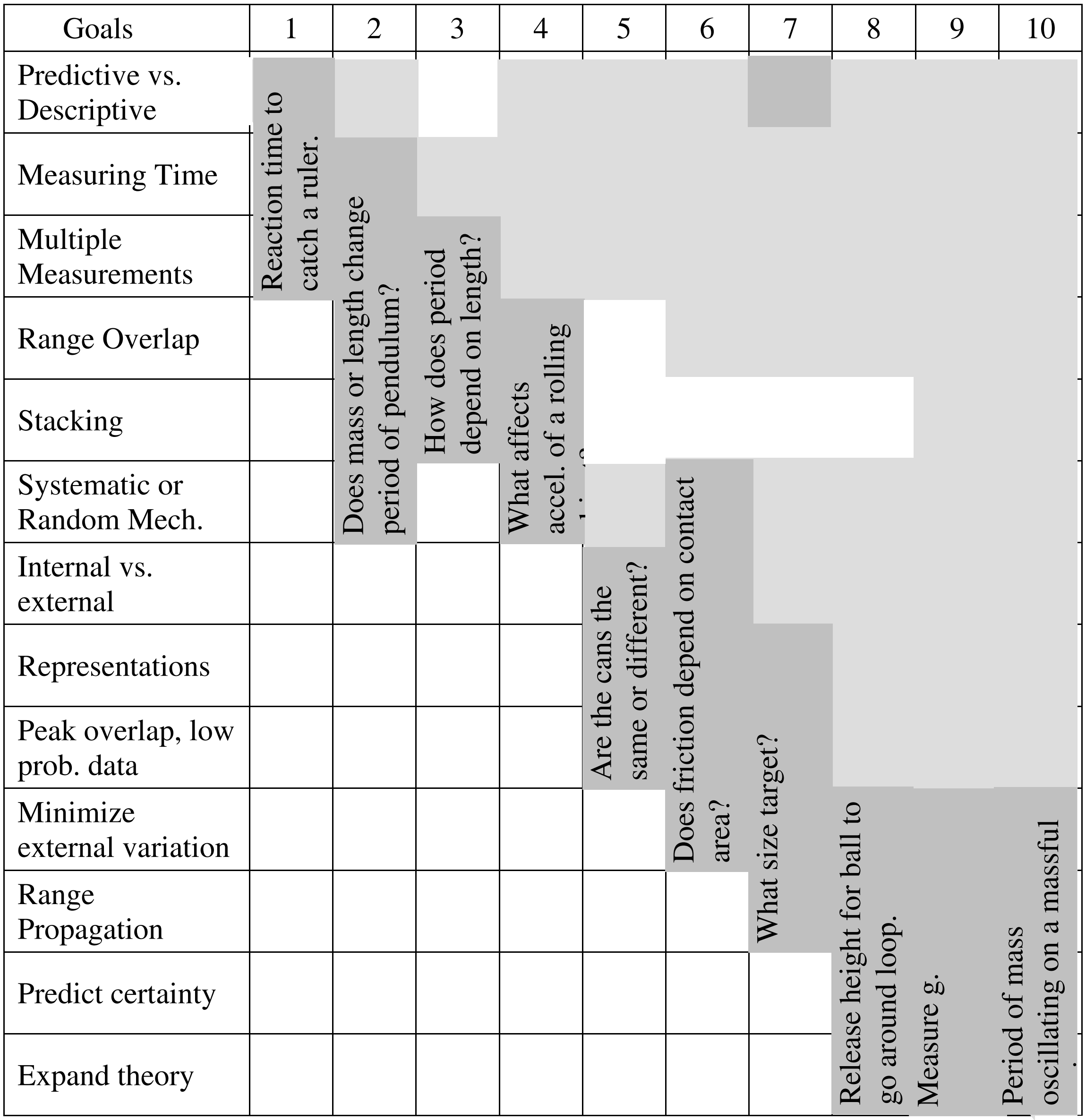


Figure 3–5

Each lab's measurement goals. Primary goals are shown with a darker fill than revisited goals.

---

[1] The exception to this occurs in labs 5 and 8, where one point from the persuasiveness section is based on success. We found that students were disengaging and needed more of a commitment to their answer.

Quizzes and Homework

Throughout the semester there are homework questions about laboratory issues mixed in with the lecture homework. This connects lab to lecture, and declares that the concepts learned in lab are an important part of the course. Along with the lab quizzes, the homework questions are a chance for students to be graded right or wrong on their reasoning.[1] Because of the community aspect of lab, students may decide that uncertainty analysis is very subjective, depending only on your classmates' arbitrary opinion. The feedback they receive through grading comments and the solutions help communicate the goal of the lab. Student's take the quiz individually, to encourage them to actively participate and build understanding during lab instead of sitting back and relying on their group. The lab quizzes are open-book and open-note, giving a reason for students to write clear weekly logs, and to take notes for themselves.

Building Community

Students in the lab must be able to communicate with each other in order to build a community. At the most basic level, this means they must have time to talk to each other and they need to be able to understand each other's words. The tools and terms board and class discussion meet these aims in addition to other goals.

Tools and Terms Board

The tools and terms board is a large flip-pad hung on the wall with a few pages designated for each section's use. It was developed originally to help groups communicate with one another during the class discussion. One group would use the word range to mean the scatter size of the data, subtracting the smallest data point from the largest. Another group used range to mean the whole possibility of results, not only what one actually measured. Without the board, neither group realized that the other meant something different by the word, and they were unable to communicate with each other. With the board, a student or TA can request that a word be put on the board and that the class decide on a definition. Sometimes the TA identifies an important concept students are using, and asks students to think up a word and definition for it. This forces the class to communicate about what the word means, and also establishes that it can and should be used. Students use the defined terms in their weekly logs and also in the lab quiz. The terms help organize and keep track of the progression of ideas in the lab – students often look to the board for possible solutions to problems they face in lab. As an example, one section defined concepts similar to accuracy and precision, calling them constant shift and inclusive spread. (See Figure 3–6.) When students define concepts which already have a common name, they use their own name but are told the common name the following week.

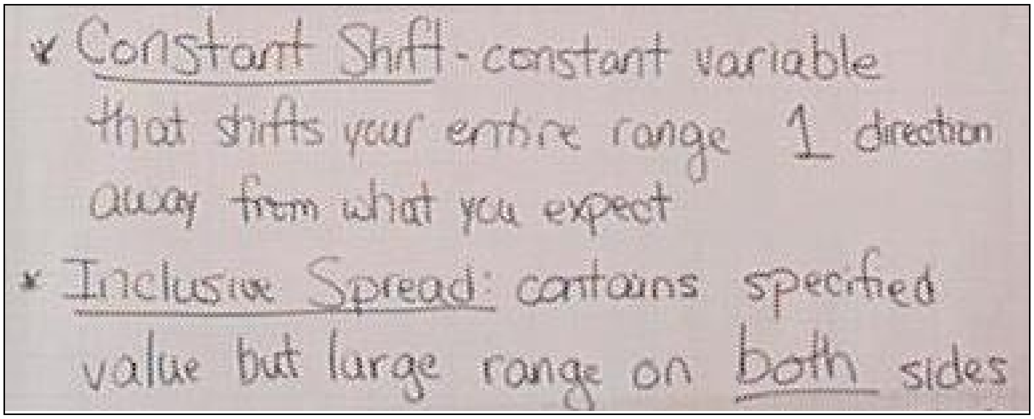
Class Discussion

During the class discussion each group stands around a whiteboard they have prepared and presents their experimental method, data, and conclusions. The goal of the presentation is to convince the other students of one's conclusion or raise problems and possible solutions if one was unable to come to a conclusion. After each presentation, the rest of the students are given time to ask clarification questions and then critique the design. If a lab is designed properly, there are several different methods to compare, conflicting conclusions to resolve, and measurement concepts to develop further. The TA acts as a director: encouraging questions, asking for clarification and elaboration, pointing out conflicts and asking for resolution, and raising possible problems. TAs are a kind of meta-discussant, evaluating the quality of the discussion and working to keep it headed in

---

[1] Both the lab homework and a lab quiz are included in Appendix C.

a productive direction. This is an extremely difficult task, even for experienced TAs, so we developed a TA training program to help them develop the necessary skills.



✓ Constant Shift - constant variable that shifts your entire range 1 direction away from what you expect

✓ Inclusive Spread: contains specified value but large range on both sides

Figure 3–6

Part of a tools and terms board with a section's definitions for concepts similar to accuracy and precision.

Preparing Instructors

The first day of class plays a large role in setting students' expectations. This makes it very important for TAs to understand the goals of lab and methods for meeting those goals before the class begins. This is done with pre-service training.

TA pre-service training

TA training begins with an introduction to the goals and structure of lab. The different aspects of the laboratory are explained and reasons given for their design.[1] There are two basic tasks the TAs must undertake in the lab: interacting with small groups and interacting with the whole class. For each of these we started by watching a short video clip of a situation and talking about possible ways to react. We also discussed the underlying diagnosis upon which each intervention was based. Different theoretical frames lead the TA to intervene differently. For example, a TA interacting with a small group of students could decide they are stuck on the concept of energy and need help with physics concepts. He may instead decide they need to make more use of the mathematical tools they have, and remind them of equations they could use. Perhaps he thinks the group is having problems communicating, so he guides a discussion between the students to clarify the problem. For every possible diagnosis we listed specific interventions. This is difficult enough to do with plenty of time and the ability to watch a video repeatedly; however, it's extremely challenging to do as a teacher in real time. To give the TAs a chance to improve at this, we instituted weekly video-watching meetings which we called video parties.

TA video parties

Once a week over lunch the TAs met to watch and discuss a short clip of a TA interacting with students. This involved first listing all possible diagnoses and interventions, and then evaluating them as to which might be more productive at reaching the goals of the lab. Frequently this involved also clarifying the lab goals. The students have a hard time understanding the new goals

---

[1] See Appendix B for the handouts given to the TAs.

of the scientific community labs and the TAs do also. This practical discussion helped illustrate the goals and allowed TAs to resolve their own framing issues.

Weekly training meeting

The purpose of this meeting was to specifically prepare the TAs for the next week's lab. It began with an evaluation of last week's lab: the experimental methods students designed, the terms they defined, the class discussion. Then the TA's were led through the beginning of next week's lab as if they were students. When the lab design was difficult, they sometimes took data, but often did not. The model of the lab introduction gives TAs a rough idea of the goals of the particular lab and sets the context. It also gives them a chance to think of questions they may have as instructors and as students. Then the TAs explicitly discuss the measurement goals introduced and revisited in this lab, ways to meet those goals, terms to be defined, and possible problems and solutions.

Grading

Grading is critical for setting the frame and giving feedback to students. It also has a large influence on student motivation and affect. One positive result of group logs is that the TAs can spend more time on each log, reading it carefully and giving thoughtful feedback. The week before an assignment needed grading for the first time (weekly logs or lab quizzes) the TAs would individually grade and then discuss two actual responses. This resulted in more uniform grade distributions because all the TAs had to agree on how many points to give for what. But it also allowed TAs to talk about how to best give feedback to students, in a discussion paralleling the video discussions described previously.

Summary

Any interpretation of results presented in this dissertation depends on many features of the data source including the physics course, student population, and laboratory curriculum. The physics course has been modified to not just teach physics concepts, but teach students how to learn science, by developing a learning environment that fosters meta-learning. The student population consists mainly of biology majors in their junior and senior year. These students are not confident in their math skills, yet are quite sophisticated in other areas such as group work and presentation skills. In interviews, students in the meta-learning course mentioned that this course required a lot of thinking, something which is rare in other courses. They were willing to do this thinking as long as they saw the benefit of it and realized that physics might help them now or in the future.

Scientific community labs are designed, as the name suggests, to form a scientific community. The students work in groups of four and must present their conclusion to the rest of the class and critique other students' presentations. The lab task is an open-ended research question: open-ended in terms of method (different methods are productive) and in terms of results (a 'true' answer is not known). A scientific community is necessary so students have colleagues to convince of their results, creating a frame where the tools of uncertainty analysis are useful. Each research question legitimately raises new measurement concepts and revisits previous concepts. The main goal of the lab is for students to build an understanding of these concepts. In the next chapter we document the student response to these labs.

References

Burnstein, R., and Lederman, L.M. (2001). Using wireless keypads in lecture classes. *The Physics Teacher, 39,* 8-11.

Campbell, D. T., and Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Company.

Duggan, S., and Gott, R. (2002). What sort of science education do we really need?. *Int. J. Sci. Educ., 24*(7), 661-679.

Ericsson, K. A., and Simon, H.A. (Revised Edition). (1996). *Protocol Analysis: Verbal reports as data*. Cambridge, MA: The MIT press.

McDermott, L. C., Shaffer, P.S., and the Physics Education Group. (2002). *Tutorials in Introductory Physics*. New Jersey: Prentice Hall.

Poulis, J., Massen, D., Robens, E., and Gilbert, M. (1998). Physics lecturing with audience paced feedback. *Am. J. Phys., 66*(5), 439-441.

Shapiro, J. A. (1997). Electronic student response found feasible in large science lecture hall. *Journal of College Science Teaching, 26*(6), 408-412.

Sokoloff, D. R., and Thornton, R.K. (1997). Using interactive lecture demonstrations to create an active learning environment. *The Physics Teacher, 35,* 340-346.

Sokoloff, D. R., and Thornton, R.K. (2001). *Interactive Lecture Demonstrations*. New York: John Wiley and Sons.

Chapter 4: Students' Responses to the Laboratory

Introduction

One basic rule of education is that the instructor needs to know the current state of her students to be able to teach and hopefully cause a positive change. In the beginning part of this chapter, we look at students in the algebra-based physics course and probe their expectations for and responses to the traditional laboratory. Through interviews and surveys we investigate students' opinions about the purpose of lab, the usefulness of lab, and uncertainty analysis. We then switch to students in the scientific community labs, analyzing their ideas about the purpose and usefulness of lab, and their complaints about lab.

Students enter the lab expecting it to be designed to meet a certain purpose and they interpret all the activities as aiming for that purpose. They think the lab useful or useless depending on whether the goal is worthy and on whether that goal is met. This makes it important to know what the students think the goal is and to communicate the proper goal to the students. We use student responses to determine the students' frame in the traditional lab and in the scientific community lab. We then compare the two frames to determine whether the scientific community laboratory design has been successful in changing students' frames to be more productive for learning measurement and uncertainty.

Students' responses to the traditional laboratory

The typical student in the algebra-based introductory physics course has experience with many different lab courses. They enter the room the first day with certain expectations about lab ranging from the responsibilities of the instructor to the arrangement of the room. These expectations will guide the student's actions and how the student interprets the actions of others. The student may be aware of these expectations, may become aware of them as they are violated, or they may remain tacit. These students also, like most students, have opinions about what is helpful and what is not, and how they would design the lab. The researcher may be able to infer student's tacit expectations from such opinions and from student's actions during lab. We describe the first in this chapter, and the second in the following chapter. We collect student's opinions and their overt expectations by interviewing the students.

The interview design

Most interviews took an hour and allowed students to freely share their ideas about the course. Three sets of semi-structured interviews were performed, all with volunteer subjects. Set A took place halfway through the first semester mechanics course and sets B and C took place after the second semester electricity and magnetism course. (See Table 4-1.) Each set of interviews had its own protocol, and the full versions are included in Appendix D. The most relevant questions are as follows:

- What is the purpose of the labs?
- Why is the lab required?
- In what way do the labs help you learn physics?
- Think of a lab you found useful. How did it help?

Set	Date	Time	# Students
A	11/00	Mid semester mechanics	10
B	5/01	Post semester E+M	13
C	5/02	Post semester E+M	6

Table 4-1

The date and time interviews took place and how many students participated

Results – Purpose of Lab

When answering these questions, students frequently distinguished between what they saw as the ostensible purpose of the physics laboratory (typically the answer to the first two questions) and what, if anything, they actually got out of the lab. In a few cases the two did coincide.

Ostensible purpose

One student, code name Bill[1], clearly distinguished the designed purpose of the lab from how he actually benefited.

> "Well, I can see their purpose, they're to show you that what the professor is saying is really true, that it really happens that way, but I believe him, I don't need the labs to prove to me that it really happens." (Bill, set A)

Students who hold this common belief, that labs are supposed to prove the theory, tend to think that labs are useless. They do not need to waste all this time because they already believe what the professor is saying.

> "Yeah, it's just annoying, and like I said before, the derivations, I just totally, the teacher tells me this is how it is, I accept it, I don't need to go to lab and prove it" (Beach, set A)

Besides causing students to think lab is useless, this belief can cause other problems. When the lab "doesn't work" because of friction, heat loss, or other real world effects, according to the student the lab fails to even achieve its useless purpose. These effects are all considered to be error, in the common use of the word. This leads students to blame the lab designer, TA, or broken equipment for their data, and to avoid any attempt to make sense of what happened. Students will not consider investigating what actually happened (a major element of real research) because that will not help further the goal of proving the theory.

This belief may originate from years of reading the first line of a lab handout which states the aim, such as "to prove that momentum is conserved in a collision." Students are failing to separate the aim from the pedagogical purpose, and this may harm their ability to learn in the laboratory. Students who are able to identify the pedagogical purpose have been shown to create more links between separate activities, to develop a deeper understanding, and to feel a greater sense of satisfaction (Hart et al., 2000).

To physically demonstrate concepts was another purpose identified by students. Instead of having the professor show a demonstration in lecture, we make the students perform it in lab. This

---

[1] For more information on Bill or any other interviewed students see Appendix D.

way they get to hear about it and also see it in action which will hopefully increase their understanding.

> "I guess ostensibly the purpose is to sort of expand on some of the principles that are presented in lecture, give you a hands on type of learning about these principles, I guess." (Henry, Set B)

A problem occurs if students do not know the concept of which they are supposed to be building an understanding. If a lab is supposed to demonstrate induction and the students have no idea what induction is, they will have a hard time seeing it demonstrated.

> "I think it's to, just demonstrate the concepts that you're learning in class. On that note, it was really hard because my lab was on Monday morning, so we didn't learn stuff until after the lab and I never really understood the labs. But, I guess it's to show, in an active setting, what you're learning in class." (Lisa, set B)

These students are distinguishing between the aim and the pedagogical purpose of the lab. But they have limited the pedagogical purpose to seeing something they already know, instead of building concepts from their observations. This will also limit students' attempts at sense-making. There is no reason for them to try to make sense of a demonstration about a concept of which they are unsure.

Actual Benefit

Some students who thought of lab as an active demonstration did find it useful.

> "I feel like I do kind of understand lab because it is a little more of a hands on, real life example I guess." (Liz, set A)

Since the student actually performed the demonstration in lab it helps her remember it better than just seeing the professor do a demonstration in lecture.

> "It's more interesting when you can see it and I'm a visual person. It helps me to see stuff, like okay I get it. I'll remember something we do in lab much more than I remember something we talk about in lecture." (Clara, set C)

One problem with relying upon a demonstration to help one's memory is that sometimes one remembers the wrong outcome of the demonstration, especially if the result contradicted one's intuition. Also, we run into the previously mentioned problem of a lab "failing" to show the right phenomenon. Labs do take place in the world, and frequently other effects overwhelm the theory the lab is supposed to be demonstrating. If the lab purpose is demonstration then such a lab not only has been useless, it also may be deemed harmful because it showed students the wrong thing.

Other students distinguished between memorizing something and actually learning.[1] A demonstration lab would only help with memorization, but these students saw lab as giving them a chance to take a concept and struggle with it, reason though it. This process helped them actually learn.

---

[1] See students' comments on thinking in chapter 3 for another example of such a distinction being made.

> "I think trial and error and hands-on discovery is one of the most important ways to learn, because then it really becomes ingrained in you. I mean someone can tell you something and make you memorize it but that's not really learning, I think it's more when you have to reason through a process like we do on our exams. …And then when you see it actually moving and why, what's going through, what's happening, and why things are moving in directions I think it was really helpful."
> (Sally, set B)

Such students took an active role in the laboratory, not only manipulating the apparatus, but also predicting what would happen, troubleshooting, and figuring out why. The laboratory gave them a chance to test out their ideas and get feedback immediately.

> "It was an immediate feedback, if it worked, something happened, and if it didn't, something else happened, and you knew that you had to change some cables around and then once you did change the cables around you were like oh so this is why, okay so connect to this and this…"
> (Sarah, set B)

If what these students report is actually the case, that they spend lab in active inquiry trying to make sense of phenomenon, then they are doing what the lab designers would like. Unfortunately, even students who distinguish between memorization and learning do not always attempt to make sense during lab, perhaps because they are in a frame where they do not see sense-making as productive[1].

Uncertainty analysis

No students brought up uncertainty analysis as a goal of lab. During the interview, students usually had to be asked directly about uncertainty analysis, except for one student who mentioned that he was glad his TA had eliminated that part of the lab. Depending on what they deemed the purpose of physics laboratories, students saw uncertainty analysis either as a necessary part of lab or as a useless waste of time. Students interpreted uncertainty analysis as a method for dealing with one's results being amiss.

> "I expect error. God only knows I'm not perfect, I've talked about being wrong enough on homework…Error is error, if I make very careful measurements I can reduce the error but I can't get rid of it…It's important to be careful, but if I'm off by 10 or 20 percent I'm not all that concerned, it's when I'm off by hundreds or thousands of a percent, the I'm like 'oh, what did I do here, why is that so bad.'" (Thomas, set B)

This student is describing more of a troubleshooting approach than uncertainty analysis, using a cutoff of 10 percent as what is acceptable and then searching for the problem. Part of this idea is productive, in that it causes the student to search for sources of external variability[2], but students who are comfortable with a 10 percent cutoff are less likely to see the need for range and data set comparison. Most students claimed an understanding of either human error (reading a dial

---

[1] See the case study of Veronica and Carl in Chapter 5.

[2] See the scientific community lab goals in Chapter 3.

incorrectly) or percent error (the difference between your result and the accepted value) but expressed confusion about anything further.

> "That's one thing that I don't really understand that well, the error analysis. [I understand] the actual, like, reasons for error, writing them down. But finding the numbers for error, I don't understand how you can find numbers for error. We did percent error in high school, but I'm not quite sure if that's what we're doing now." (Mark, set A)

This lack of understanding contributed to students' dislike of error analysis. They frequently classified it busy work, or claimed it did not belong in a physics course.

> "I can calculate error but I don't want to… I would rather just have to do the critical stuff than actually have to calculate the other stuff the statistical stuff….I don't like to do the extra stuff that's not applicable towards the class. If I want to take a statistics course, I'll take it and do statistical analysis later." (Arnold, set B)

The belief that uncertainty analysis is extra and not critical to the laboratory is extremely detrimental, especially for the scientific community labs which have uncertainty analysis as a main goal. This means that the scientific community labs must first convince students that uncertainty analysis (in the scientific sense) is a useful goal, that it is in fact the pedagogical goal of the labs, and then teach students how to do it. We can investigate whether this actually happened by probing students' responses to these labs.

Students' responses to the scientific community laboratory

Students' responses to the scientific community labs were gathered by unstructured written feedback and through an anonymous survey. We use these results to determine whether the students are in a frame which is productive for learning measurement and uncertainty.

Unstructured written feedback

During lecture a few weeks into the semester the professor asked the students to take out a sheet of paper and write down any feedback they would like to give us anonymously about the course. This was repeated at the end of the course. We gathered 146 feedback responses in the middle of the semester and 77 responses at the end. The number of responses was greater for the first round because students were much more frustrated and motivated to give us feedback early in the semester. Attendance to lecture remained constant at about 150 students out of 160 throughout the semester.

Students offered feedback about all aspects of the course, including the professor's hair. Out of all of the responses, 76 students (52%) commented on the lab in the middle of the semester, and 44 (57%) at the end. The lab responses were coded, resulting in the following seven major categories, in order of decreasing frequency: labs are too rushed; the grading is too harsh; the design of labs is good; labs are useless; the purpose is unclear; labs are useful; and, labs are stressful. The number of responses in each category divided by the total number of lab responses yielded the percent of responses from that category. For example, in the middle of the semester about ten percent of the lab responses contained a comment that the labs were useful. This means that ten percent of the students who mentioned the laboratory found the labs useful, which is seven students. Since some responses fit into more than one category, the sum of all the percentages is greater than 100. The standard error bars (square root of N divided by total number of lab responses) in Figure 4–1 demonstrate that only the first four categories showed a significant change from mid to post semester.

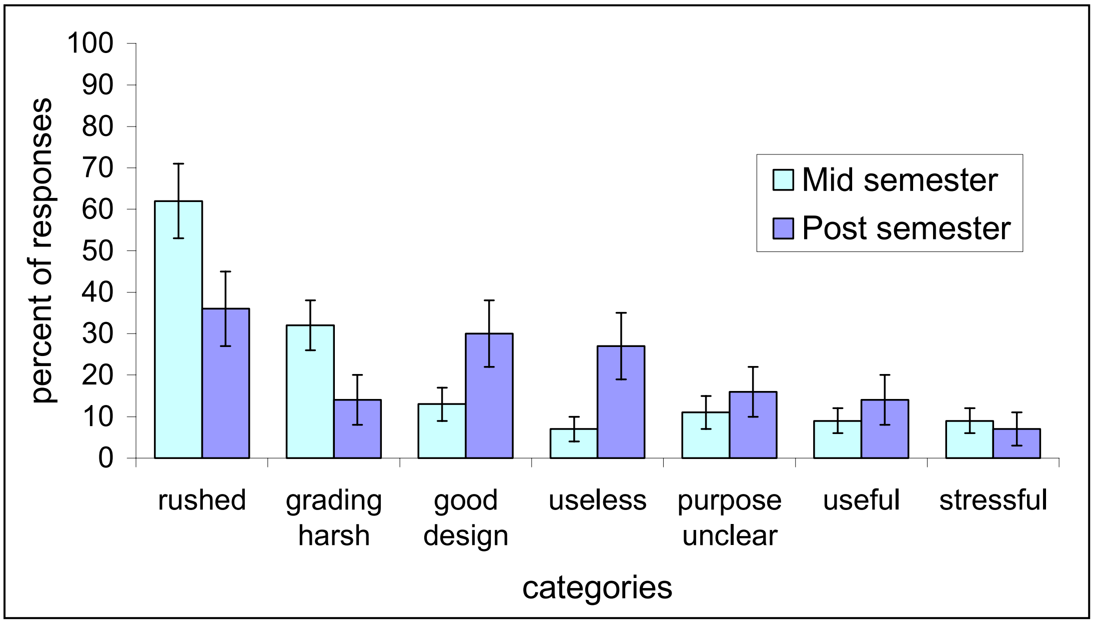


Figure 4–1

Percent of student responses from the mid and post semester unstructured written feedback

Labs are too rushed

Lack of time was a recurring problem in the scientific community labs. Because of the class discussion, students had around 50 minutes to plan, execute, and write up their experiment. This was enough time if students quickly thought of a plan and were efficient taking data. During several labs toward the end of the semester, a few groups in every section would finish early and then wait for class discussion. However, if something went wrong causing them to revise their plan, or if the group had a hard time coming to a consensus, then they would run out of time. The mid semester feedback occurred directly after the third lab, a mathematical modeling pendulum lab. This lab particularly required too much of the students, and students were justified in complaining about lack of time for that lab. As the semester progressed students become more efficient setting up equipment and taking data. These are likely the reasons for the decrease in the percent of students complaining about labs being too rushed.

Grading is too harsh

Students were required to hand in their group report at the end of class, so they were worried that the lack of time would harm their grade.

> “Considering the time restraints, we should at least be given some kind of method or ideas from the TA’s so that we can get everything done in time or, the other alternative is to take into account the enormous time constraints when grading labs. However, this was obviously never done, as everyone in my section got a 5 or less [out of 10] on the lab.”

The grading of the group report was not based on successful completion of the lab. It was possible for groups without a conclusion to get a perfect score as long as their evaluation section was well-developed with concrete suggestions for improvement. Unfortunately, students did not seem to understand or perhaps believe this because it was so different from what they had previously experienced. There were fewer complaints about grading at the end of the semester, perhaps because the grading rubric became clearer, perhaps because students' reports improved and so their grades improved. The grading rubric was one of many aspects of the scientific community labs which clashed with student's expectations, making clear communication (by written and spoken word and actions) imperative.

Another major clash occurred during the class discussion. At the end of lab each group was given time to present their experiment and convince the class of their conclusions. This was designed to mimic a graduate student presenting her research to the rest of her research group. In such an environment, the graduate student might show some data and present her conclusions and then ask for help or other opinions on certain aspects. While presenting her conclusions she would hope to be challenged, and given a chance to defend her research and perhaps refine her conclusion. If she presented to a room full of silent people who just stood up and left after she finished, she would feel that something was drastically wrong with her research. This would be true even if people said "good job" as they stood up and left.

In the scientific community lab, students viewed their presentation as failing if anyone pointed out problems with their research. They would also consider it a failure to have to ask the rest of the class for any help on their research. If anyone pointed out problems, the presenters would defend themselves by responding "well, we didn't have enough time." Frequently TA's tried to intervene by noting that the group could write the critique in their evaluation section of the report, in an attempt to convince the class that critiquing is a good thing to do. However, many students were not convinced.

> "when we presented to the class, [the TA] told us it was wrong and had the class point out our errors. How is this teaching us anything? I still don't know if I did it right and I hate to ask questions, lest I be met with another response that makes me feel stupid."

This student did not see the point of the class discussion; he did not think it was teaching anything. Even worse, it made him feel stupid, which means he will be less likely to work hard at the next lab, for fear he will fail again. We tried to structure the class discussion as a time for the whole class to come together and use each group's work to build the strongest conclusion. But students were in a classroom, with all the expectations of years of being a student. In the typical classroom frame they are expected to be right. Being wrong, especially in front of a large group, not only makes one look stupid but may lower one's grade, and is to be avoided at all costs. This frame was very difficult to change.

Labs are useless

This complaint may be related to what students regard as the goal of lab. The scientific community labs were not designed to demonstrate theory; in fact, they frequently disproved an oversimplified theory[1]. If students think the goal of lab is to demonstrate theory, then it is not surprising if they find the lab useless. These labs were designed to build an understanding of

---

[1] In lab 11, for example, students consider the equation for the period of a hanging object oscillating on a 'massless' spring. They are asked how we could update this equation for use when the mass of the spring is comparable to the mass of the hanging object.

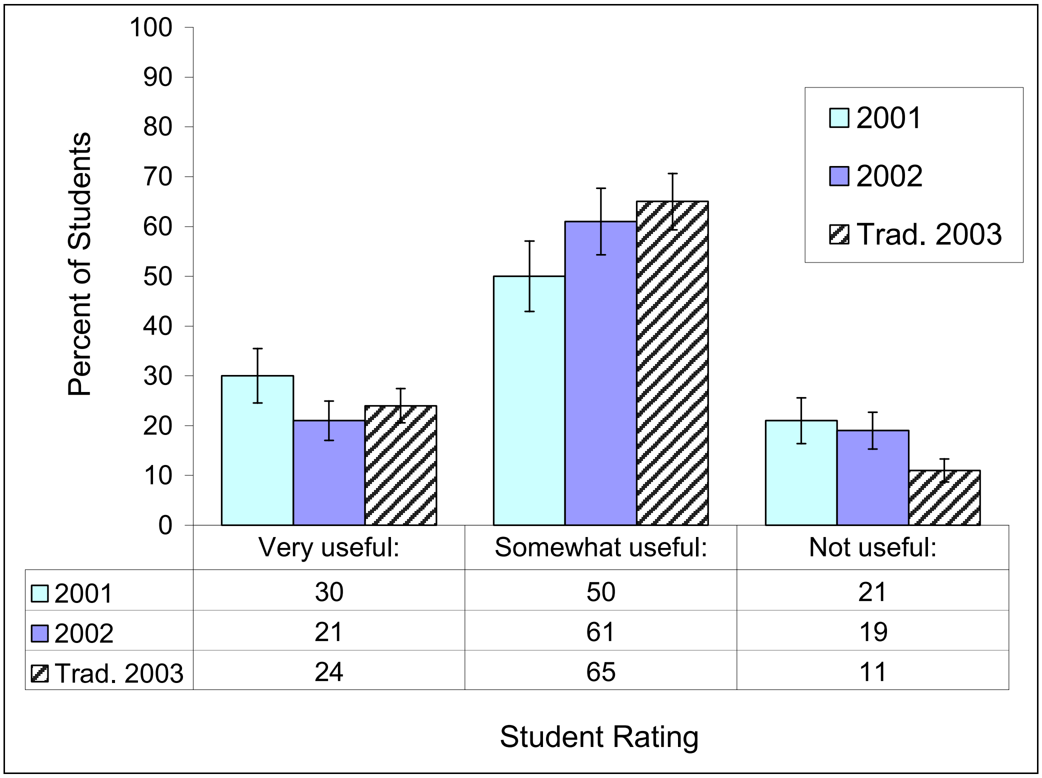
measurement and uncertainty. Perhaps students either did not understand these goals or did not think them worthwhile. In an attempt to disentangle these two issues, this problem was investigated further through an anonymous survey given at the end of the semester.

Anonymous Survey Results

A short anonymous survey[1] was developed to probe what students thought was the purpose of the labs and how useful they found the labs. It was administered at the end of the scientific community labs in the fall semester of 2001 (in part, to 101 students) and 2002 (in whole, to 135 students). For comparison data we surveyed 203 students from the traditional "cookbook" mechanics labs in spring 2003.

Usefulness rating

We asked students "How valuable were labs in helping you learn physics, compared to labs in a typical science course?" and required them to choose very useful, somewhat useful, or not useful. The percentages and error bars were calculated as for the written feedback coding above. Around one quarter of students found the labs very useful or not useful and about a half found them somewhat useful. (See Figure 4–2.) The only difference between the traditional labs and the scientific community labs is the fewer number of not useful responses for the traditional labs. This question specifically asked students about learning physics, which is not a goal of the scientific community labs and is the main goal of the traditional labs. Even so, students judged both lab designs to have about the same value for learning physics.



	Very useful:	Somewhat useful:	Not useful:
□ 2001	30	50	21
■ 2002	21	61	19
▨ Trad. 2003	24	65	11

Figure 4–2

Students' rating of the usefulness of the scientific community labs

---

[1] See Appendix C for the exact surveys given.

Between 2001 and 2002 we made an attempt to clarify the goals of the lab for the TAs and the students. Specifically, each lab handout for the TA contained the pedagogical purpose of that lab and we encouraged the TA to write these on the board at the beginning of lab. In the first lab, the TA's led a class discussion about the purpose of the lab. However, this did not change students' usefulness rating. To find out what exactly students think is or is not useful, we need to know what they see as the purpose of lab.

Perceived purpose of the scientific community lab

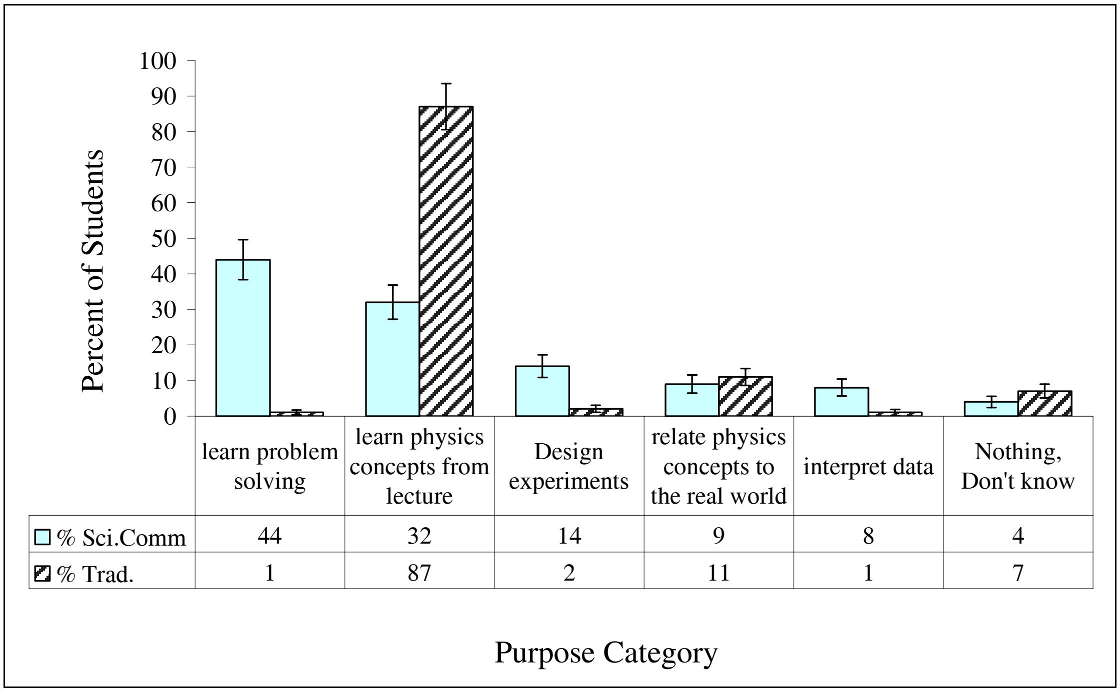
During the 2002 scientific community labs we asked students a free response question about the purpose of lab. The wording of the question for the scientific community labs, "What was our (Prof. Redish and TA's) main purpose in lab?" was chosen to ask for the pedagogical purpose instead of the aim of the lab. Unfortunately, a few students interpreted that to mean our specific actions during lab, responding "to ask us questions and make sure we were thinking." These responses (15 out of 140 students) were disregarded. The remaining responses were coded into the following five main categories listed in decreasing frequency:

- Learn problem solving, thinking, and creativity
- Learn the physics concepts from lecture
- Learn how to design an experiment
- See how physics concepts relate to the real world
- Learn how to interpret data

The main goals of the scientific community labs, learn how to design an experiment and learn how to interpret data, were identified by only 14% and 8% of the students, respectively. (See Figure 4–3.) Almost a third of the students thought the purpose of the lab was to build on the physics concepts they had learned in lecture and almost half wrote to learn problem solving, thinking, and/or creativity. Neither was a stated goal of the labs, but students often had to use these skills while working in the lab.

For the traditional class, we changed the wording of the question to "What was the lab designer's main purpose for the labs?" More than 85% of the students in the traditional lab stated that the purpose of lab was to build on the physics concepts learned in lecture. Some students used this question as a chance for sarcasm, responding "to keep you busy for at least ½ hour" and "to take up my valueable [sic] time". Only three students mentioned problem solving or thinking skills and only eight mentioned experimental skills.

The failure of so many students in the scientific community labs to identify the main purpose of the lab is rather disappointing, but we need to interpret this result in light of the traditional lab environment so familiar to the students. In the interviews, all of the traditional students mentioned some form of physics conceptual learning as the goal of labs, and in the survey 87% of the students mentioned this. One could imagine students saying "what else could the purpose be?" Only 11 of the 203 traditional students mentioned any kind of skills like problem solving or experimental design. Students in the scientific community labs have realized there is more to learning physics than pure conceptual theory. In contrast to the traditional students, none of the students in the scientific community labs thought the purpose of lab was to prove a theory. One student specifically mentioned this, saying the purpose of lab was "to get us to come up with conclusions by ourselves instead of just doing an experiment to 'prove' them." Students in the scientific community labs do have a more productive sense of the purpose of lab, but there is plenty of room for improvement.



	learn problem solving	learn physics concepts from lecture	Design experiments	relate physics concepts to the real world	interpret data	Nothing, Don't know
☐ % Sci.Comm	44	32	14	9	8	4
▨ % Trad.	1	87	2	11	1	7

Figure 4–3

The purpose of scientific community and traditional labs as identified by students

Future Use

Students in the scientific community labs also mentioned problem solving or experimental design skills in their answer to the free response question "What did you get out of lab that will help you in your future career?" In the scientific community labs, the previous categories of problem solving, designing experiments, and data analysis were repeated. Students also thought that skills of working in groups, debating and evaluating, and time management would be helpful. Only five percent of students mentioned data analysis. We have either failed to convince the majority of students that data analysis is useful or they felt that they didn't learn any data analysis in the labs. Only sixteen percent of students claimed that nothing from the scientific community labs will be useful, compared to 38% in the traditional labs. Even though few students mentioned the main goal of the scientific community labs (data analysis), they recognized other useful skills.

No students in the scientific community labs responded that physics concepts will be useful, as compared to 24% of students in the traditional labs. Students in the traditional labs also mentioned that learning to follow directions, such as a laboratory protocol, will be useful to them. Clearly no students in the scientific community labs mentioned this, because they did not have any steps to follow. The idea that learning to follow steps should be a goal of the physics laboratory is an interesting one; it is true that many students go on to work in a laboratory and will have to follow a protocol. However, we believe that it is extremely important to understand why the protocol is designed a certain way, because people who do not understand the basis for the protocol will be more likely to ignore crucial elements to save effort, and will be unable to judge if the protocol was designed effectively for slightly different situations. It requires an understanding of measurement and uncertainty to truly follow a protocol.[1]

---

[1] See the section on applicability to biology in Chapter 3 for a discussion of the measurement concepts related to a protocol on water quality.

Summary

We used interviews and surveys to probe students' ideas about the purpose of lab and the usefulness of lab. Students in a traditional laboratory viewed the purpose of lab as to prove theory, to demonstrate concepts, or to allow active inquiry about concepts. They did not mention any form of data analysis or other skills as a purpose. In the scientific community labs, the majority of students saw building an understanding of physics concepts as the main purpose. Unlike traditional students, they also mentioned several skills they learned, such as problem solving and creative thinking, that would be useful for the future.

Unfortunately, the main goal of the labs, measurement and uncertainty analysis, was mentioned by relatively few students. When students enter a classroom with certain expectations, it is very difficult to change them. It has proven especially difficult to cause a large change such as that from 'proving physics theories' to 'learning how to analyze data'.

The information reported by students seems to indicate that many students are still acting from a traditional classroom frame, where students are expected to find the right answer and give it back to the teacher when asked, all the while competing with other students to be the best at this task. This caused students to worry about time pressure, believing that they were graded on having a complete correct answer. This also made it difficult for students to critique each other and view critiques as useful.[1] Students worried that any negative comments may indicate that they did not complete the task and may lower their grade instead of seeing them as helping the whole class move toward a better solution. Having described student's self-reports, in the next chapter we investigate students' behavior during lab to determine the students' frame.

References

Hart, C. M., P. Berry, A. Loughran, J. Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments? *Journal of Research in Science Teaching, 37*(7), 655-675.

---

[1] Likely there are many other causes for students' reluctance to criticize and be subject to criticism. We discuss this further in Chapter 9 as an area for future research.

Chapter 5: Social Construction of an Understanding of Measurement and Uncertainty

Introduction

"My students are just not thinking" is a common complaint by instructors, especially after a laboratory class. Many instructors have amusing stories to tell about a student who defended an answer that clearly was ridiculous, or who wasted time, hurt someone, or broke equipment by doing something in lab which the instructor considered foolish. In a University of Maryland traditional introductory physics lab on heat and work, two students spent 15 minutes measuring out 40 mL of water very carefully, only to weigh the container empty and full to calculate the volume of water. Instead of making sense of their method, such students appear to be just following instructions. They are not trying to make sense of the experimental method and are not monitoring their actions by keeping track of what they have done and what they need to do to make progress toward their goal.

Introductory Physics Laboratory activities are typically very different from other components of a class (lecture and recitation): they last longer, are self-paced, and require students to work in groups to complete a task. Since students control who does what at what time, they need to be metacognitive about their activities during lab, planning and evaluating their progress toward the goal. Students must be in a frame where they view such sense-making and metacognition as productive activities, and the lab design should encourage such a frame.

In this chapter we present a method for evaluating the amount of sense-making a group of students perform during class. A large or small amount of sense-making by a group of students suggests that the students are in a cognitive frame where they consider sense-making productive or unproductive. The design of the laboratory contributes toward setting the frame, so this is one way to evaluate the lab design. In addition, if students do not know what their goal is, what they are currently doing, what they should be doing next, and how that will help them reach their goal, they are less likely to benefit from lab. By counting the number of productive metacognitive statements we evaluate how much students are monitoring their actions, and then evaluate whether or not their metacognition is productive. We use this method as one way of comparing the traditional and the scientific community laboratory designs.

Cognitive Behavior Mode

We analyze student's verbal communication during the laboratory, using a discourse analysis tool to infer students' cognitive behavior from their verbal behavior. There exist many different discourse analysis studies on the laboratory (for example, Keys, 1996; Kelly and Crawford, 1996; and Roth, 1994). The data source, the size of the group of people being analyzed, and the size of the utterance being coded are all designed to help answer the specific research question. The particular attributes of the coding scheme presented here are designed to answer the (very broad) question "Are students thinking during lab?" Of course, students are always thinking, but not always thinking about productive topics in productive ways.[1] We want students to be using their

---

[1] For a definition of thinking according to the students, see students' responses to introductory physics in Chapter 3.

existing knowledge to actively build new knowledge relevant to the goals of the laboratory. We define this as *thinking*, or more specifically, *sense-making*. This transforms the question into "Are students sense-making during lab?" In later sections we clarify and expand the goals of the coding scheme, but this main question is useful for conveying the overall purpose of the coding scheme.

Data Source

The question "Are students sense-making during lab" requires that we study student's actions during lab, and furthermore, that we identify actions that give evidence for sense-making. All introductory physics laboratories at the University of Maryland require students to work in groups, causing students to talk to each other. These interactions are what we analyze for evidence of sense-making. Data is taken from videotapes of the student group (2-5 students) working together during lab. The videos are transcribed and coded using both the transcript and the video. Traditional cookbook laboratories frequently began with a 5-15 minute lecture given by the TA. This was not coded, as it was not a group interaction, and rarely did any students speak. The scientific community labs ended with a 30 minute whole class discussion, which also was not coded, because it was TA led and very different from the small group interactions.

Group Mode

The data we choose to analyze are videotapes and transcripts of group interactions. One reason for this relates back to a refinement of the main question. Do we want to know if individual students are sense-making, or if the whole group is sense-making together? Consider a situation where one student shows evidence of sense-making and the other group members do not. In such a case, that one student would be thinking out loud and there would be no interplay with the rest of the group. Our students being fairly competent in their social ability, such a student would quickly stop talking, and, if they continued sense-making, would do it silently. In fact, most one-sided thinking never got that far: a student would make a request for sense-making, and would let it pass if the rest of the group refused to answer their request. The transcript we first introduced in Chapter 2 gives a good example of this.

1 *C*: Acceleration equals change in velocity over change in time.
2 *B*: But then velocity, what's delta v?
3 *A*: How did we do this last week? Go back to last week's.
4 *B*: I didn't understand last week's.
5 *A*: But we just need to find it. How did we do that whole thing with the final velocity?

In lines 2 and 4 student B makes a request for understanding, which she drops when the rest of the group denies her request. The lack of evidence for one-sided thinking makes coding individuals fruitless. Such a request could have caused the rest of the group to try to make sense of acceleration and velocity, and then the majority behavior of the group would have been sense-making.

Another motivation for coding the group's behavior relates to the main goal of the scientific community laboratory. The labs are designed to create more authentic scientific interactions between students so that students will construct an understanding of uncertainty to build and evaluate convincing arguments. Analyzing group interactions documents how an individual's cognitive development can occur in the context of a group interaction. A third reason is simple feasibility: coding individuals is often impossible, as they may spend large periods of time silent or just agreeing with the speaker. We are more interested in, and have more data on the group's behavior, and so we code the group's majority behavior.

Coding Sense-Making, Logistics, and Off-Task Behavior

Our coding scheme classifies student behavior into three cognitive behavior modes: Sense-making, Logistics, and Off-Task. In the following sections we define each of these modes and describe how to identify them.

Off-Task

Of the three modes, off-task is the easiest to identify. Students who are off-task are talking about something unrelated to the laboratory. They know that it is unrelated, and will often return to the task when a TA walks by. During one off-task conversation a student asked the rest of her group to help her decide if her boyfriend was cheating on her, and if so, what she should do. The students carefully weighed the evidence given and offered their suggestions. One could code their thinking in this instance as making sense of human behavior; however, we limit the code of sense-making to areas relevant to the physics lab. Non-relevant behavior is coded as off-task whether it is sense-making or not.

Logistics

During the laboratory, students frequently have to be doing activities such as reading, writing, taking data, gathering equipment, performing calculations, and reporting to the laboratory teaching assistant. While these are all necessary parts of the laboratory, during them students are not explicitly constructing new physics knowledge. We code these activities as logistics. The following is an example of a discussion about timers which was coded as logistics.

1: Do any of you have a watch with a stopper on it?
2: They have timing thingies there.
1: The timers here are really bad. Like, they just don't work. You click on them, they just do whatever they want.
3: You can use this if you prefer.

Frequently a group in the logistics mode appears to have little continuity in their conversation because the group members are busy performing tasks and are not responding to the previous person's statement.

1: 6.62
1: 6.63
2: Oh, perfect.
1: Excellent.
3: What different lengths do we need, we need, 10?
2: Oh wait, this is wrong, that's the wrong. It's different string. There's two rolls.

Sense-Making

The sense-making mode was defined by Otero (2001) as including "discussions or utterances that were based on real or hypothetical experiences, peer instruction, and discussions or utterances about inconsistencies or unresolved issues." Among other things, students can be making sense of formulas, physics concepts, their experimental design, or their data. While in the sense-making mode, students may or may not be correct or successful, but they are using resources that are frequently productive for learning science.

Consider the following conversation, where students are attempting to design a method to determine whether two identical appearing cylinders are the same or different on the inside.[1]

> *2:* Let me just, let me play with my idea here. [rolls cylinders] Did you see compared to this one how this one rolled consistent and this one kind of like jolted?
> *1:* I saw it rolled faster
> *4:* I didn't see anything
> *2:* Well watch again and compare. I mean, I'm not saying this is scientific or anything like that. See how this goes slow at first and then speeds up?
> *1:* Yeah, I did see that.
> *2:* That's what I'm trying to say.
> *1:* But it's hard for me to tell whether you gave it a little jab when you let it go.
> *2:* Well, let me do it this way. Can I borrow your pen?
> *4:* I don't think that makes a difference.
> *2:* Because it [the question on the lab handout] says whether the insides of the cylinders are the same or different.
> *4:* But I think what he's asking is are they the same mass. Like, I think what we're supposed to be looking for is mass.
> *2:* Because you can have the same masses on the inside but they can be distributed differently. You know what I mean?
> *4:* But I don't see, how is that testable? What experiment would you use to test that?

The students are attempting to make sense of the question asking if "the insides of the cylinders are the same or different", as well as trying to design a method to answer this question, so this conversation was coded as sense-making.

There are several characteristics of the sense-making conversation that are different from the logistics conversation. One of them is greater continuity in the conversation. Statements are clear responses to previous statements, students ask questions of each other, those questions are answered, and the conversation is all on the same topic. Students also give reasons for their statements, using information from previous experience, lecture, the lab handout, data, rough observations, or instincts. Because they are talking about ideas and are disagreeing with each other, students are more likely to qualify their statements by saying such things as "I'm not sure", "maybe", and "I think". Students also are making some progress and not just repeating the same problem or idea. They may not solve the main problem, but they will have solved smaller problems along the way. These qualities: continuity, reasons, qualifiers, and progress, do not define sense-making. They are not required for a conversation to be sense-making, and they sometimes occur in logistics. However, a conversation with a large amount of these qualities is likely to be sense-making.[2]

Again, note that a group's behavior is classified according to the mode they are in: sense-making, logistics, or off-task. We code students' conversations in chunks and not their single

---

[1] This lab took place several weeks before rolling was considered in lecture.

[2] See Sandifer (2001) for more on sense-making discussions: a classification of sense-making statements, and an identification of factors providing support for sense-making discussion.

utterances. For example, line 2 on page 62 could be coded singly as sense-making, yet taken in context it would not be so coded.

Student behavior in different lab designs

Our coding scheme is useful for comparing students' behavior in different lab designs. Here we compare the traditional cookbook laboratory in introductory algebra-based physics at the University of Maryland to the scientific community labs. In cookbook laboratories students are given a handout that describes the equipment they are to use and the steps they are to follow. In the scientific community labs, students are given a paragraph setting up the research question and then an outline of the structure of the period.[1] They are expected to interpret the question, design and perform their experiment, and analyze their data as a group with help from the TA and other groups.

Two different coders analyzed four of the seven labs presented here, with an inter-rater reliability above 85%. After discussion, the two codings agreed by more than 95%. Sense-making may not seem to be well defined, but it is reliably recognized.

Results of the coding show that students spend more time sense-making in the scientific community labs (see Table 5-1), so this lab design successfully creates an environment that encourages sense-making. Students are also spending less time off-task, likely because the scientific community labs feel more intense[2], so the students do not have the time or inclination to go off-task.

Lab Design	% Sense-Making	% Logistics	% Off-Task	Time (min)
Cookbook A	4	85	11	94
Cookbook B	4	82	14	62
Scientific Community A	20	78	2	77
Scientific Community B	21	77	2	66
Scientific Community C	22	78	0	49

Table 5-1

Student behavior in cookbook and scientific community labs

Students' use of metacognition

Most of the time during both lab designs is spent on logistics. This is not necessarily wasted time – as mentioned before, logistics is necessary. However, students engaged in logistics should be monitoring their progress toward the goal, and not just performing unrelated tasks. To study the amount of monitoring students engaged in during a lab we coded for explicit metacognitive statements.

Metacognitive Coding

Metacognition is defined roughly as thinking about thinking, or "knowledge and cognition about cognitive phenomena" (Flavell, 1971, p. 277). It is any cognitive process that refers to or

---

[1] The design of the scientific community labs is described in detail in Chapter 3. The specific lab handouts are contained in Appendix A..

[2] In traditional Cookbook Labs at the University of Maryland, students often finish the lab early and leave before the end of the period. This does not happen in the Scientific Community Labs, partly because of the time taken by the whole class discussion at the end of lab.

controls any part of cognition, both knowledge about cognition (knowing that one is likely to forget a phone number and so should write it down) and cognitive regulation (detecting an error in one's reasoning). The following examples should help clarify this definition.

One of the most common statements coded as metacognitive is of students evaluating their own understanding. It is usually in the negative sense, "I don't get this", but may be in the positive, "This lab makes so much sense to me, more than other labs." Students may evaluate the reasoning of themselves or their lab partner. In the following example, one student is converting g*cm to kg*m, and another student evaluates her reasoning (statements coded as metacognitive are labeled **M**).

1: Cm to m is 2 decimal places, and g to kg is three, so I divided this by one hundred thousand.
M *2:* That might not be right, though.

Students may evaluate their data, their method, or their actions. They also may plan what they need to do next, "What do we do now?"

In contrast to sense-making, logistics, and off-task, metacognition is not a group behavior mode. It is performed at one time by an individual while the group is in any cognitive mode. An exception to this occurred during the scientific community lab C, where three students in a group began planning what they could do to make next week's lab go easier.

M *4:* Yeah, next time I'm really gonna sit down and think about how we're gonna do it.
1: I know, so I know what to do: 'wait, I have the plan'.
2: But it's hard like, because we didn't know what materials we would have coming into this, you know, you have to kind of wait so you can see
1: Yeah, that is the problem, you don't know
2: Cause you could make up this wonderful thing and not

All of these statements proposed or evaluated the cognitive activity of planning, but we code only the first as metacognitive, because that is what began the conversation.

Metacognition is much more difficult to code reliably than behavior mode. In one lab, both coders identified 24 metacognitive statements, but only 12 of those were the same statements. In another lab one coder identified 56, another coder 45, and 17 agreed. However, there were a total of 545 statements in this lab which could have been coded metacognitive, and the two coders agreed that 478 statements should not be metacognitive. This is an agreement greater than 87% on the not-metacognitive. Because of the problem identifying metacognition, only those statements which both coders first identified as metacognitive are taken as metacognitive.

Metacognition in different lab designs

To compare the amount of metacognition from students in different lab designs, we count the total number of metacognitive statements and divide it by the total time to find the number of metacognitive statements per hour. Surprisingly, both lab designs had significant amounts of metacognition (see Table 5-2). Unfortunately, this large amount of metacognition does not seem to have any effect on student's behavior. There is no correlation between amount of sense-making and amount of metacognition.

Lab Design	meta/hr
Cookbook A	8
Cookbook B	15
Scientific Community A	12
Scientific Community B	18
Scientific Community C	21

Table 5-2

The number of metacognitive statements made per hour for different lab designs.

One reason for this is that the coding scheme considers all metacognitive statements as equal. In the following example we show three statements, both coded as metacognitive, but actually very different. The first two metacognitive statements (labeled **M**) come from cookbook lab A, on the topics of work and heat. The last comes from the end of scientific community lab A where the students were asked to distinguish between identical-appearing cylinders and had planned their method using the idea that the different cylinders had liquid or solid contents.

[near beginning of lab]
2: What's the calorim-eeehh thing?
1: It's a calorimeter.
M *2:* Ok. I really don't know how this is going to work.
1: Oohh, we're supposed to…[reading manual]
2: We have to weigh it, though. [goes to do that]

[second time running experiment]
1: It's exactly the same. 135.
M *2:* That's so weird. They're going to think we copied it. But we didn't copy it.

M *3:* No, I guess I was in liquid solid mindset, which completely destroyed what
I needed to think about….we could have examined the data without notions
of liquid and solid in our heads, it would have helped us be more objective.

In the first conversation, student 2 admits that he does not understand the lab, but his admission does nothing to change the students' behavior. They continue reading from the lab manual and following the written steps. In the second conversation, students realize that the second trial of the experiment yielded the same result as the first. The students recognize that this is odd, but they worry only about what the TA will think. In the third conversation, the student recognizes that her model of the cylinders affected her method and her data analysis, and caused her to come to the wrong conclusion. She gives a suggestion as to how to avoid this problem in the future, but does not get a chance to try it, as this is the end of the lab. The first two metacognitive statements feel very different from the third, yet the coding does not distinguish them. One way of distinguishing productive from unproductive metacognition is to see where it takes place during the lab and whether it causes a change in mode. We do this with the representation described below.

Cognitive Behavior Representation

To represent the mode and metacognitive statements of a group of students, we use a time-line representation derived from Schoenfeld (1985). Time spent in each mode is shown as a bar, and metacognitive statements are identified by inverted triangles on top of the bar (See Figure 5-1). This allows one to see the time evolution of student's behavior and to analyze the causes for transitions between modes. Metacognition which is productive is circled.

Productive Metacognition

The time-line for Scientific Community Lab B (see Figure 5-1) shows the first transition from logistics to sense-making occurring around 22 minutes. During this lab the students were trying to figure out if two identical-appearing cylinders had the same configuration of mass inside. Until this time the students have been individually trying to weigh the cylinders in their hands, asking the TA for help, and attempting to understand the lab question. Student 1 activates sense-making mode by asking a clarifying planning question, which is coded as metacognitive. This question causes the rest of the group to start to work together to figure out a procedure.

M *1:* What's our planned procedure?
3: We could find the acceleration
2: You don't know the mass.
3: You don't need the mass to find the average acceleration.
2: Isn't it?
3: No, if we just do velocity over time, and then we get the acceleration, and then, we don't know the mass, but we could find out the force, well, if we race them, are they going to roll at the same speed?

There were several metacognitive statements before this one that failed to activate sense-making. These statements were unproductive, perhaps because the rest of the group was too confused to go into sense-making or because the metacognitive statement was not a good activator. Another productive metacognitive statement took place around 51 minutes. The students were racing the two cans down a ramp to see which can would win. Student 1 admits confusion and asks the group for an explanation. This sends the group into sense-making mode.

3: [Can] 11 still wins. [Can] Nine can't keep up with us.
M *1:* I'm just confused why it was tied before.
3: It's probably the way you were releasing it.
4: If you see the instant replay [the starting gate] was actually pushing [the cylinder] a little bit.

By admitting confusion, student 1 causes the group to move from taking data to making sense of their data and proposing causes for it.

There are cases where metacognition is productive in highlighting a problem but does not cause a change in mode because solving that problem does not require one. For example, in scientific community lab C students are measuring the period of a pendulum hanging off a support taped to the desk. One student mentions that the support should be moved farther out because the pendulum is going to hit the table legs. Another student moves the support out, and the conversation moves on. Productive but non-mode-changing metacognition is not identified by this scheme, as the scheme focuses on the relationship between behavior mode and metacognition. In addition, metacognition during sense-making cannot be judged productive by this definition (unless the group decides they are doing too much sense-making and should transition into logistics).

In summary, we have found that the scientific community lab design is associated with more sense-making behavior. Students in both labs were coded as having a lot of metacognition;

however, the scientific community lab design resulted in more productive metacognition in students, represented by the circled triangles. (See Figures 5-1 through 5-6.) The scheme described in this chapter is useful for evaluating the amount of sense-making and productive metacognition during any lab design, and it also may be used for case-studies, to investigate different students' interactions with the lab design. We perform a case-study in the next section.

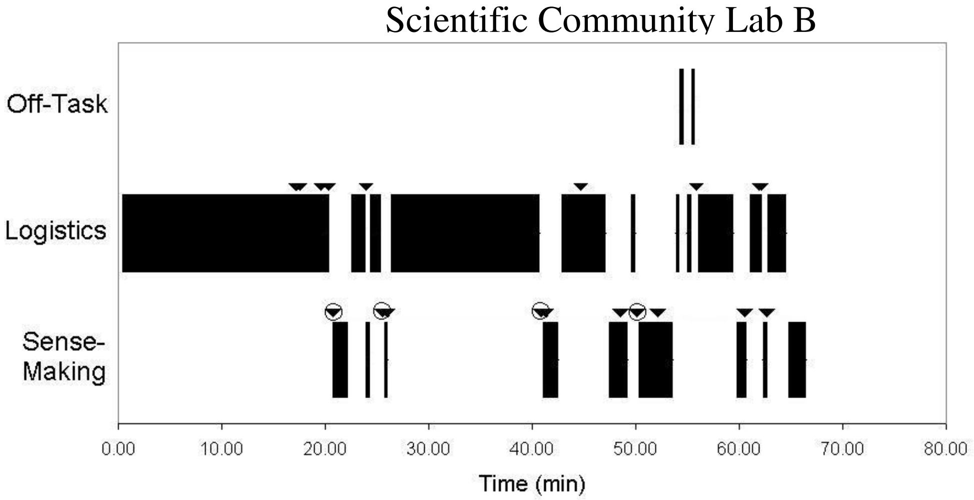


Figure 5-1

Cognitive Behavior Time-line for Scientific Community lab B

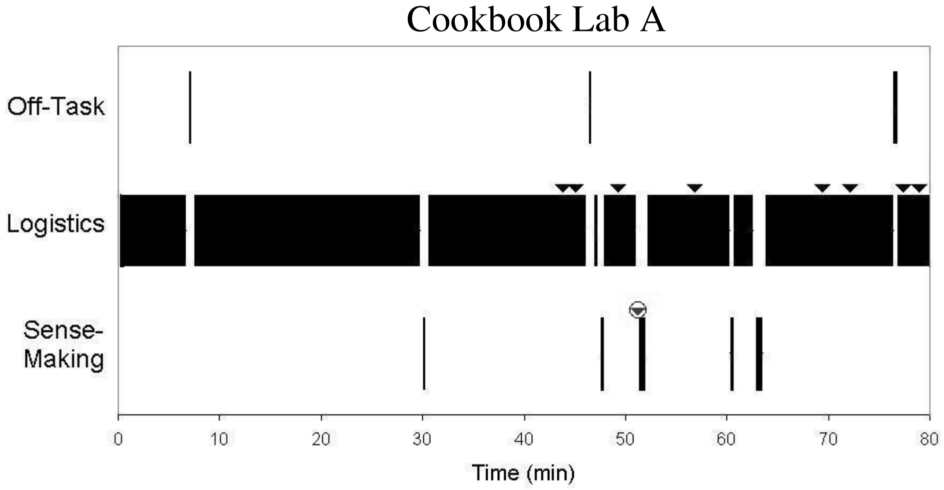


Figure 5-2

Cognitive Behavior Time-line for Cookbook lab A

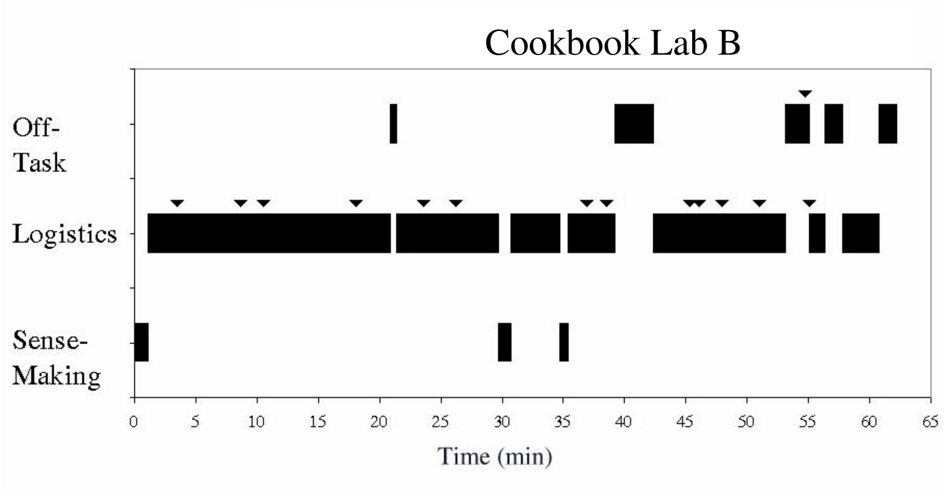


Figure 5-3

Cognitive Behavior Time-line for Cookbook lab B

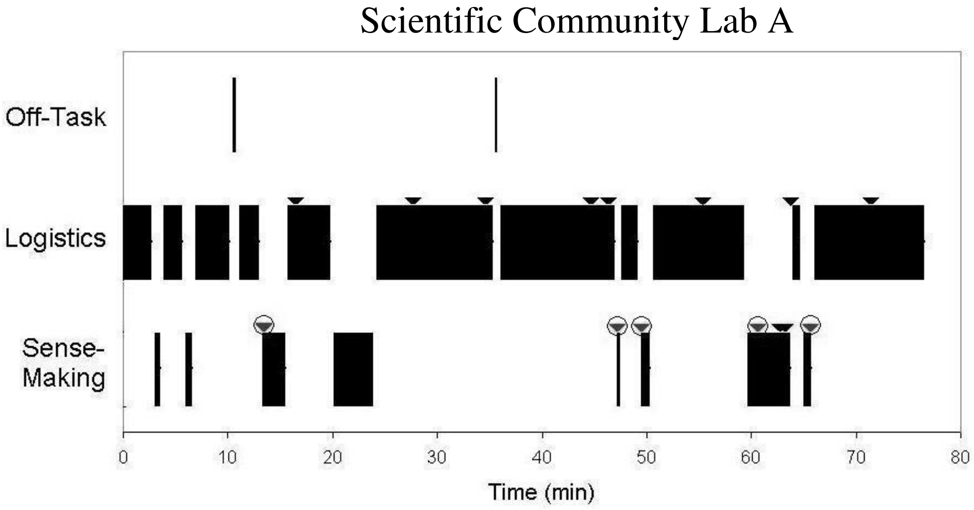


Figure 5-4

Cognitive Behavior Time-line for Scientific Community lab A

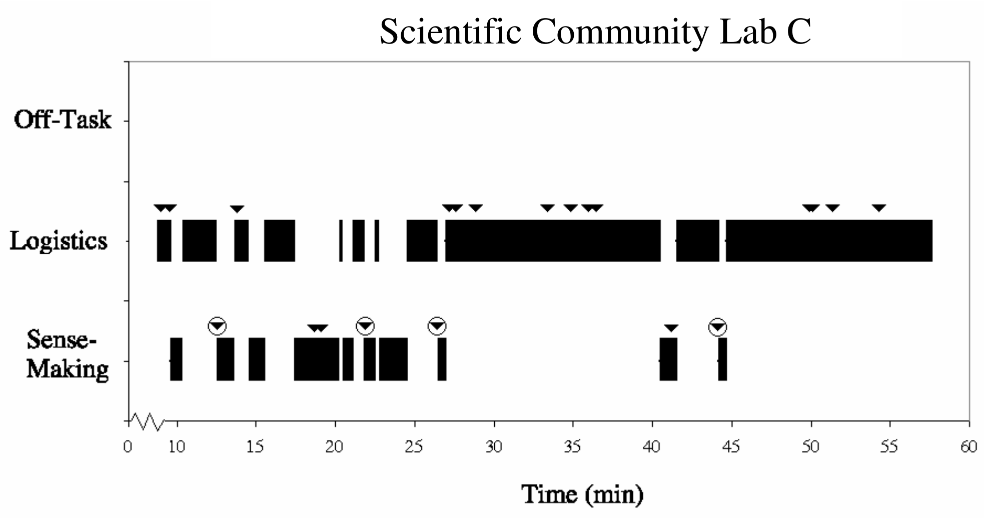


Figure 5-5

Cognitive Behavior Time-line for Scientific Community lab C

Case Study

In this section we use the behavior mode coding scheme to investigate students working on another type of lab, *Cookbook+Explanation* labs, taught during a traditional section of the algebra-based physics course at the University of Maryland. I had been assigned to teach a cookbook laboratory and was only allowed to make small changes. I created Cookbook+explanation (*C+E*) labs[1] by inserting 'explain why this happens' types of questions into the lab manual and cutting some of what I deemed 'busy work'. The group of two students[2] analyzed working through the C+E labs is among the best group in the class and demonstrates the most productive response to this type of lab. Because of the atypical group, this is a case study of this group's interactions with the lab design, and not an evaluation of the lab design. We videotaped and coded this group working on two C+E labs, the seventh and ninth labs of the semester.

If we compare the amount of time spent sense-making in C+E labs, it is surprising the difference caused by the small change in lab design. (See Table 5-3 and Table 5-1.)

Lab Design	% Sense-making	% Logistics	% Off-task	Meta/hr
Cookbook+Expl A	17	77	6	12
Cookbook+Expl B	38	62	0	6

Table 5-3

Student behavior in cookbook+explanation labs

In C+E lab A students used a magnet to induce current in a coil of wire connected to a galvanometer. The students, Veronica and Carl[3], spend all their time in the logistics or off-task mode until they come to the modified section of the lab. Here the TA asked them to first predict the

---

[1] The lab handout for Cookbook+Explanation Lab B is included in Appendix A.

[2] Students worked in groups of two in the Cookbook and Cookbook+Explanation labs, and in groups of four in the Scientific Community Labs.

[3] Veronica and Carl are same-gender code names.

direction of the induced current and then explain how the current was being induced, writing "a paragraph as if you were explaining this to someone who didn't understand this but was in physics." This causes them to transition into sense-making (see Figure 5-6, around 32 minutes). After struggling to make sense of the phenomena and failing, they take a break for a bit in off-task. They then return to the same question, and succeed in making sense.

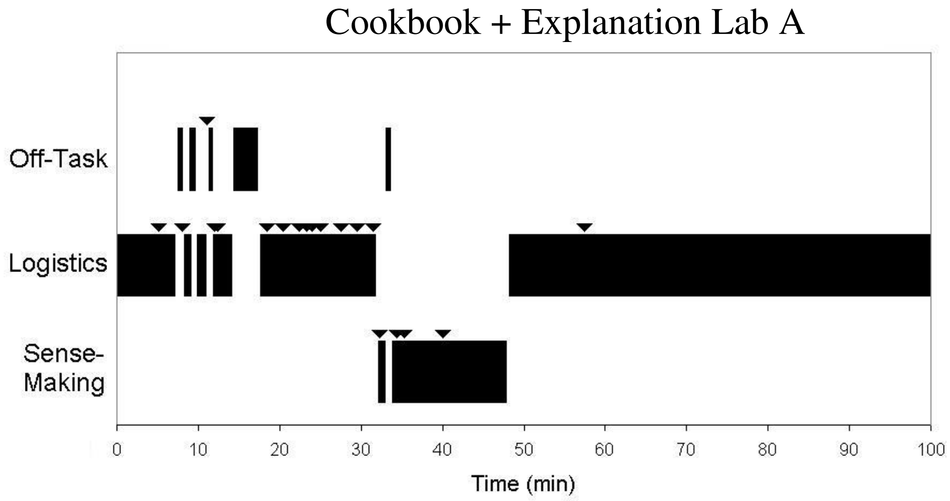


Figure 5-6

Cognitive Behavior Time-line for Cookbook + Explanation lab A

C+E Lab B provided students with a laser, screen, and slits of varying widths and separations to study diffraction.[1] For this lab, the TA rewrote the whole manual, requiring students to explain their reasoning after every new observation. These questions apparently forced Veronica and Carl to go into sense-making mode more frequently (see Figure 5-7).

Transitions to sense-making coincide with the lab questions "explain why this happens," "predict what it will look like," and "do these trends make sense". This indicates that the rewritten lab manual set the frame for Veronica to believe sense-making was appropriate and useful in this situation, and so she went into sense-making mode much more than in previous labs. She noticed this herself, and commented in the middle of the lab "I like this lab. This lab makes so much sense to me, more than other labs." Because Veronica was an excellent student it was very easy for her to go into a sense-making mode, (it just took rewriting the lab manual). Also, once she was trying to make sense, she was likely to be successful.

---

[1] The transcript for C+E lab B is included in Appendix F.

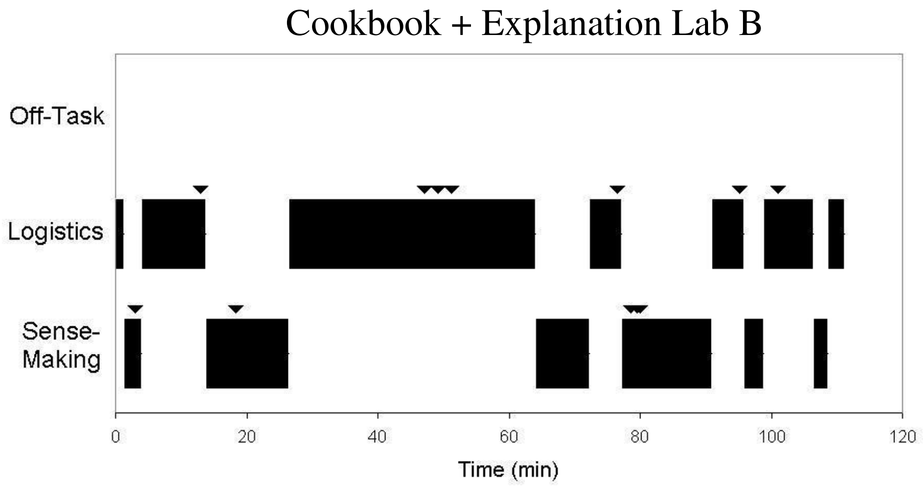


Figure 5-7

Cognitive Behavior Time-line for Cookbook + Explanation lab B

In other class activities, Veronica frequently makes sense of physics using her real world experience. During one activity another student in her group thought common sense was not productive to use in that context, and instead was trying to make it "physics oriented." Veronica responded, "It is, it is physics-oriented. That's just the way it is," and convinced her group to use the common sense explanation. However, in the cookbook lab setting, she only attempts to make sense when she is forced to by a specific question. In fact, for their first transition in C+E lab B it is Carl (C) who points out the "explain why" question in the lab manual, causing Veronica (V) to start sense-making:

> *C:* So the angle theta decreases, right?
> *V:* I don't know….
> *C:* Why? **She [TA] said why** is the space between the spots rising. [Points at lab manual]
> *V:* Between each one, it's just there's more light coming through for a wider one
> so more light rays interfere….
> [Emphasis added]

For many of the other students in this lab this change did not activate sense-making. Instead they used their old strategies of lab survival: ask the TA, ask other students, and read more in the manual. These old strategies can be seen by looking at the questions asked to Veronica and Carl by the lab group in front of them. During C+E lab B, these questions all concerned low level logistical information, such as "For number three, what kind of picture does she [TA] want?" and "For the second question, are they just asking for when you do one slit?". The lab group in front of Veronica and Carl was unable to make sense, and instead was struggling to follow steps.

Summary

In this chapter we introduce a coding scheme used to classify the cognitive behavior of a group. This coding scheme identifies three modes of behavior: sense-making, logistics, and off-task, and marks metacognitive statements made by students while in any of those modes. Metacognitive statements, in general, abound in the laboratory. By analyzing when the metacognitive statement occurs we can see if it caused the group to change modes, defining this as productive metacognition. This scheme is useful for evaluating the design of a laboratory to see if it creates an environment which encourages sense-making and productive metacognition. We use it to compare the cookbook and scientific community lab designs, and show that the later encourages more sense-making and productive metacognition.

We also use the coding scheme to do a case study of two students working on two cookbook+explanation labs. We show that these students were guided completely by the lab manual, and only made sense when they were asked. This is in contrast to other contexts, where one of the students is frequently seen making sense from her own volition. It seems that her frame in the laboratory causes her to view sense-making as something which must be externally regulated, and not something she should initiate.

The results of the coding scheme show that students in the scientific community labs appear to be in a frame where sense-making and metacognition are viewed as productive activities. This will hopefully help students to use the right resources to build an understanding of the concepts underlying measurement and uncertainty. In the next chapter we analyze lab videotapes to determine some of these underlying concepts.

References

Flavell, J. H. (1971). First discussant's comments: What is memory development the development of?. *Human Development, 14,* 272-278.

Kelly, G. J., Crawford, T. (1996). Students' interaction with computer representations: analysis of discourse in laboratory groups. *Journal of Research in Science Teaching, 33*(7), 693-707.

Keys, C. W. (1996). Writing collaborative laboratory reports in ninth grade science: three case studies of social interactions. *School Science and Mathematics, 96,* 178-186.

Otero, V. K. (2001). *The Process of learning about static electricity and the role of the computer simulator.* Unpublished Doctoral Dissertation, San Diego State University and University of California, San Diego.

Roth, W. M. (1994). Experimenting in a constructivist high school physics laboratory. *Journal of Research in Science Teaching, 31*(2), 197-223.

Sandifer, C. (2001). *Factors influencing Middle School students sense-making discussions in their small-group investigations of force and motion.* Unpublished Doctoral DIssertation, University of California, San Diego, and San Diego State University.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic Press, Inc.

Chapter 6: Underlying Concepts for Understanding Measurement and Uncertainty

Introduction

Previous research on student understanding of measurement and uncertainty has identified students' ideas concerning taking and evaluating data, and furthermore has classified those ideas using different frameworks.[1] But where do those ideas come from? What concepts underlie students' decisions in the laboratory? We address these questions in this chapter.

The methods an instructor uses may depend on his underlying theoretical framework, perhaps on whether he thinks students are applying unproductive resources or whether students have misconceptions. If the former, students must learn to activate resources productive to the situation. If the latter, students' misconceptions must be eliminated and replaced with correct conceptions. That is, unless their misconceptions can be understood as having an underlying structure made up of resources or other ideas. Then instruction may consist of helping the students to rearrange those existing underlying ideas, to remove some ideas, or to activate some new ideas. Knowing the underlying structure of students' larger ideas can be useful for an instructor, giving them more information to work with when deciding on an intervention. It is also useful for the researcher, allowing a more detailed study of students and how their ideas change.

In this chapter we use the method of knowledge analysis to find students' underlying ideas. We then describe these underlying ideas and use them to analyze examples of student reasoning. We demonstrate how these ideas can help an instructor decide how to intervene during class, and also how they can inform curriculum development.

Knowledge Analysis Method

To determine the content of students' underlying ideas we use a method called *knowledge analysis*. Knowledge analysis is described in detail in Sherin's (2001) work concerning the underlying ideas students use to understand physics equations. We first describe the specifics of how we used the method, and then summarize the main ideas, benefits, and limitations of the method.

Finding and analyzing measurement nuggets

In all 10 scientific community labs during the Fall 2001 and Fall 2002 semesters we videotaped two groups in each section (a total of 28 groups) and collected lab reports as well as any relevant homework or exam questions. We also videotaped selected labs from the Spring 2002 semester, the electricity and magnetism (second semester algebra-based physics) laboratory. The videotapes were the main source of data; the others served as triangulation sources. While watching the videotapes, we tagged any sections where students were making sense of their measurement method, data analysis, or conclusion in a way that involved their data. For example, we would not tag a video clip of students using a kinematics equation to decide how they could find acceleration from a time and length measurement, but we would tag a video clip of students arguing how a short,

---

[1] See the review of previous research on students' understanding of measurement and uncertainty in Chapter 2.

steep slope would affect their data compared to a long, shallow slope. These selected video clips, which we called *measurement nuggets*, were then transcribed. Measurement nuggets were taken from groups of students working alone, groups talking to the TA, groups informally talking to other groups, or groups formally presenting their results to the rest of the class. This process led to more than 100 measurement nuggets, averaging about 15 statements each (that is, about 25 lines of transcribed text each).

The next step was to sort these nuggets into categories according to subject. Some of the nuggets involved students talking about "human error": its effects or how to reduce it. In other nuggets students argued about whether it is better to keep experimental conditions constant (such as having the same person do all the timing) or to change them (have different people timing). In still others, students reasoned about two different sets of data, considering whether the data sets disagreed or agreed. Some of the nuggets did not fit into a general category. Human behavior is very complex, and we do not claim to be able to unpack any argument students could make during lab. However, the few categories we do describe were very common, occurring in the majority of the groups working on the scientific community labs.

We then studied each category separately with the main goal of determining what underlying concepts students were using to create their arguments. We determined an analysis scheme for each, deciding which theoretical framework would be most productive for describing students reasoning, and further sorting nuggets within categories. At each iteration we would develop the analysis scheme, apply it to the measurement nuggets, and use it to re-sort the nuggets. If the analysis scheme was not reliable (two researchers did not agree on results) or could not be used to describe all of the nuggets in that category, we revised the scheme. After several iterations the different schemes coalesced into their final versions, resulting in two resource-like analyses and one (mis)conception-like analysis.

Benefits and Limitations of Knowledge Analysis

Knowledge analysis is a method for finding and describing the contents of students' underlying ideas, and it assumes that students' understanding consists of a large number of smaller pieces. It is meant to be used on reasonably large domains (such as introductory physics) which likely contain a large number of pieces. The final goal of such a study would be a complete description of all these pieces, so that one is able to describe any behavior found in this domain (Sherin, 2001). However, such a task is beyond the scope of any single project, so we must limit our domain to a reasonable size. We chose to focus on students' understanding of data: taking it, analyzing it, and building conclusions. This means that we will not be able to describe all of students' behavior, even all behavior during the lab. The research is based on a set of highly selected data, so it is probable that we missed some ideas. This study is not an extensive list of all the ideas students use to understand measurement; it is an analysis of a few of those ideas. However, these are the ideas we found most common in the data, so they are likely to be found in many other student laboratories, and should be useful both to researchers and instructors.

Measurement decision resources

Two common student ideas we identify appear to fit with the resources perspective. These ideas are not right or wrong by themselves, but can be applied in a productive or unproductive manner. We also show that they can be applied in many different contexts. We first describe these ideas with examples, and argue that these ideas are resources. We then discuss how these ideas can benefit an instructor.

Data Source

The data which contributed to this analysis came from many different labs and from different parts of those labs (the intro discussion, group work, presentations, and the ending class discussion). In fact, the first resource has been recognized in contexts other than the laboratory, such as a group of students working on homework problems or on a tutorial. This wide variety of data and contexts is part of what helps identify these ideas as resources, as opposed to the conception-like idea described in a later section.

Description of Resources

A constant has no effect

By the time they are juniors and seniors in college, the majority of students are adept in the skills of controlling variables, at least for simple experiments. In the scientific community lab on whether mass affects the period of a pendulum, every group knew to keep length (and to a lesser extent, amplitude) fixed while changing the mass of the pendulum. Keeping the length fixed allowed the students to conclude that any change in the result was due to the changing mass - to a certain extent, keeping the length constant meant that length had no effect on the result. This idea we will call *constant or no change leads to no change in result* (*a constant has no effect* for short).

This resource can be applied productively, when controlling variables, or unproductively, as an excuse not to worry about something, as in "we kept [whatever] the same so it would not affect our data." We counted 16 examples of *constant has no effect* in the set of measurement nuggets. One productive example occurred near the beginning of a lab when a group was planning how to measure the proportionality constant, k, between the acceleration of an object down a ramp and the gravitational field strength, g. One student proposed that they measure the acceleration for different ramp angles. Another student argued that this would harm their ability to determine if the shape of the rolling object was changing k.

1: I don't think we want to change the angle
2: Why?
1: Because we have no reason to.
2: Why?
1: Because it wants you to determine if the shape has an affect on k, and you want to have everything else constant.

An unproductive example occurred with a group whose setup involved a mass hanging over a pulley attached to the table. The mass was hitting the edge of the table, and student 5 pointed this out.

5: Is this supposed to be touching the end of the table?
4: That does not matter, as long as it stays the same.

In this lab the students were measuring the normal modes of string vibration to determine the mass of the hanging object. Even a constant effect would still change their final result, making their measured mass too large or too small.

Another example occurred during a radiation lab, where students used a Geiger counter to find the half-life of a radioactive source. Many groups began by measuring the background counts of the room, often for periods of one minute. During their group presentation, Group 1 mentioned that they ignored that data. In the transcript, G1 stands for a particular student from group 1, G4 is a student in group 4, TA is the teaching assistant for the section, TA2 is an extra helper for that day, and line numbers are included for easy reference.

1 G1: Also with the background radiation, we just left it because it's a constant, it's affecting everything the same amount, so why deal with it?
2 TA: Um, ok, ok, but
3 G1: That's your zero.
4 TA2: That's an issue you guys talked about [G4], you wanna?
5 G4: We did it first actually with the background in and then we did it in Excel without the background and it [the half-life] differed by about, I don't know, almost a minute and a half.
6 G1: Really?
7 TA: Ok…

Student G1 considers it obvious that the background radiation does not matter – he seems to take a stance that it is a waste of time to worry about it. In line 3, he uses more scientific terms to explain, perhaps like zeroing a scale, or defining the zero for potential energy. Then, in line 6, he is quite surprised and skeptical (this is clearer in the video) that including the background radiation made a difference in the calculation.

Part of what demonstrates that this idea is a resource comes from the fact that it can be applied in productive or unproductive ways, as just shown. Another reason is that it can be applied in many different situations. The *constant has no effect* resource could be used in social situations (little Johnny is always fidgety, it does not mean he is not paying attention), in biology (our sense of smell desensitizes to odors over time), and even in fairy tales (the boy who cried wolf). This resource is actually exploited in the field of linguistics, when studying the development of language. Researchers are able to tell how many different phonemes an infant can distinguish by slowly varying a sound and measuring when the infant reacts. When the infant keeps hearing the same sound, he keeps a constant sucking rate on a pacifier, but when the sound changes, his sucking rate briefly increases (e.g. Eimas et al., 1971). One could argue that *constant has no effect* is based on a survival instinct: people are not able to pay attention to everything, so they limit their attention to things that are different, since such things are more likely to be dangerous. This resource could have been generalized from any of these contexts.

Same is same

Suppose someone had a picture of a person lying on the beach and a picture of a person skiing. To determine if they were the same person, she might compare the height and build of the two people, the shape of their nose, eyes, and mouth, and their hair and skin tone (if visible). She might conclude that they have the same eyes, nose, and height, so they are the same person. She is applying the *common attributes means same type of thing* resource, or *same is same* for short.

Students often use the *same is same* resource in the laboratory when deciding whether two sets of data agree or disagree. The students choose the most compelling attribute, such as the average, the mode, or even the size of the range, and if those are the same the students conclude that the two sets of measurements agree. In the following example, student 1 is presenting his results to the class about whether two identical-appearing cylinders have the same or different internal structure.

1: Our averages actually turned out to be pretty exact, they were at 1.25 on both of them, pretty much exact, very very similar results …and so our results are pretty conclusive that they're both the same object.
2: Wow.
1: Our averages turned out to be 1.251 and 1.2509, something like that.

The students go on to mention other tests they ran to make sure the two cylinders were the same, and one student in the audience objects "But when the acceleration's the same, you don't even need [further tests]". In this case, it was very obvious to the students that the accelerations were the same, which implied that the cylinders were the same.

Like the *constant has no effect* resource, the *same is same* resource can be applied as a way of limiting the effort people must put forth to survive. It would be very difficult, if not impossible, to prove that two things are the same. No two things are ever identical but they can be close enough so it is productive to treat them as the same type of object. It is possible and often productive to compare one or several attributes of two objects to determine the objects' congruency.

This resource can also be used in many different contexts. Infants seem to often test this idea. (Does this object act the same when I put it in my mouth and then throw it off the highchair?) When buying computers, one might compare hard drive, RAM, processing speed, and operating system to see if two computers are equivalent. People stereotype because they decide that people with a certain attribute can be treated the same. The IRS tax laws treat all those in the same tax bracket as the same. The *same is same* idea can be used productively or unproductively in many different contexts, so we argue that it is a resource.

Instructional Implications

These two resources, *a constant has no effect*, and *same is same*, were both identified from the set of about a hundred measurement nuggets. The first was identified in 16 nuggets, and the second one in 12 nuggets. Having argued that these ideas are resources, we now give an example of how they can help inform instruction.

Scientific community lab 10 gave students the task of measuring the gravitational field strength, g, in the basement of the physics building well enough that they could tell g to 1 percent or less. In one section 2 groups chose to use a pendulum, measuring the period (T) and length (l) and then using $T = 2\pi\sqrt{\frac{l}{g}}$ to calculate g. By measuring the time for 10 or more periods, they were able to get precise but inaccurate results. The other three groups chose to drop a heavy object and time its descent, getting imprecise but accurate results. (See Figure 6-1.) The TA, knowing that the students may have decided that a constant (such as the pendulum length measurement) will not affect their results, decided to focus on this point. He drew the representation in Figure 6-1 and asked students what could be causing the differences. Students initially thought that the length measurement would not make a difference as long as each group consistently measured the length of each pendulum. A group which included the length of the mass or a group which measured to the top of the mass, they argued, should get the same result as long as they were consistent. Eventually the students determined that different length measurements, though consistent, would make consistent differences in the group's data (and that inconsistent differences would cause inconsistent results). A constant has no effect on the range size of the group's measurement, but does have an effect on the average. The *a constant has no effect* resource can be productively applied to range size and unproductively applied to average in this context. Through this discussion the students defined terms for systematic and random uncertainty.

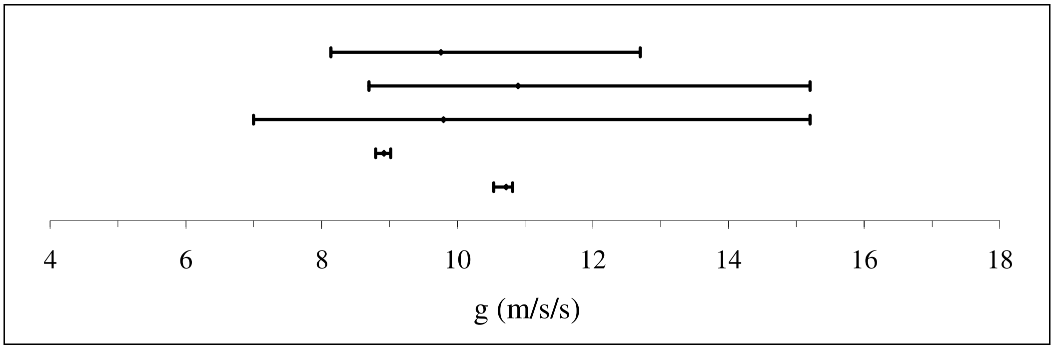


Figure 6-1

The spread in the results of five groups' measurements of g. The three upper groups timed a falling object and the two lower groups timed the period of a pendulum

Knowing that students were likely to apply *a constant has no effect* informed the TA's decision about what to focus on in the class discussion. The resource *same is same* also has instructional implications, especially when used by students as the last step in comparing two sets of data. We analyze this in more detail in the next section.

Same/Different Analysis

In the scientific community labs, four of the ten labs ask students to compare two or more sets of data to see if they are the same or different. In the Physics Measurement Questionnaire[1] (Buffler and Allie, 2001) one quarter of the questions ask students to compare two sets of data. Data set comparison is very important in uncertainty analysis, and in the scientific community labs data comparison is where students first see the usefulness of range. Student's arguments concerning data set comparison are more complicated than simply applying a resource described above; they act more like conceptions[2] with an internal structure which we describe. However, we also argue that they are not simply right or wrong: like resources, whether they are productive or unproductive depends on the context in which they are applied.

Data Source

Most of the data which contributed to this analysis came from the scientific community labs 4 and 5, which were specifically designed to address the issue of data set comparison[3]. In lab 4, students chose one property of a rolling object (mass, radius, shape, material, etc) and investigated whether that property affected the object's acceleration down a ramp. In lab 5, students were given two "cylindroids" which appear identical (except for a number) but which may or may not have different internal structure, and were asked to determine if their cylinders were the same or different (see Figure 6-2).

Lab 3 also asked the students to make a similar same/different decision; however, students' arguments in this lab tended to be less sophisticated and less clear than in labs 4 and 5 because of a lack of data, a lack of time, and a lack of understanding. The labs after labs 4 and 5 moved on to other measurement and uncertainty topics, so students' arguments about same or different were less clearly articulated.

---

[1] For more information on the physics measurement questionnaire, see the section on Buffler and Allie in Chapter 2.

[2] We do not use the prefix "mis" here because we are not making any judgments concerning whether these arguments are right or wrong.

[3] See Chapter 3 for the goals of each lab. The lab handouts are included in Appendix A and B.

You have just been allowed a marvelous chance to study the actual finches that Darwin collected. You discover, hidden at the bottom of an old box, a collection of new creatures. These ‘cylindroids’, as you call them, all have round, silver colored bodies, travel primarily by rolling, and have the same mass. There are two different species of cylindroid, but you don’t know which species a certain cylindroid belongs in.

In this lab each group of students will test two cylindroids and determine if they belong to the same species or a different species.

Figure 6-2

The lab question for Scientific Community Lab 5

Description of Representation

We use a flow-chart-like representation to describe students’ steps of reasoning from the two sets of data to their conclusion. Note that this chart is not claiming to represent what the students actually think (if such a thing exists) but rather what they use to argue their point to the other students. Students may or may not believe their argument; they may be saying what they think will be most convincing to the other students, what they think the TA wants to hear, or what another group member forced upon them. But however they came upon it, they have presented this argument to their own group members, the TA, or other groups of students, so they consider it of some worth.

We introduce this representation with an example from the *same is same* transcript described above. The students are reasoning from two sets of data, which we have called Data Set A and Data Set B (see Table 6-1). In the students’ argument (see Figure 6-3), those two data sets are represented by their average, the only information the students compare in this simple example. The averages they use are shown below the box on their respective sides. Students then perform an action (signified by a diamond shape) on this data by comparing the two numbers, and conclude that the averages are the same. Then they apply the *same is same* resource (shown by the irregular shape) to conclude that the cylinders are the same. The resource controls (or allows) the link between “same average” and “same cylinder”, so it is shown connecting to the link. To simplify things, in the rest of the arguments we analyze we will look at only the argument up to the level of the “same average” box.

# trials:		1	3	1	6	1	3	7	3	4	1	
A (sec)		1.15	1.17	1.18	1.19	1.20	1.21	1.22	1.23	1.25	1.30	
# trials:	1	4	2	1	5	1	2	4	1	5	2	2
B (sec)	1.13	1.15	1.16	1.17	1.19	1.20	1.21	1.22	1.23	1.25	1.28	1.30

Table 6-1

The frequency of each time measurement for data sets A and B

Different student arguments

The flow-chart representation allows us to classify different arguments according to their complexity: the amount of information students consider and the number of links they draw between different pieces of information. The simplest argument is that shown in Figure 6-3, where the students use only one piece of information (the average) from each data set and reason in a linear fashion, with each box having only one link into it and one link out of it. Other arguments with a similar structure compared the mode, median, or size of the range. Students may also have used a different method for comparing the data, such as percent difference. One section defined a new tool called the *Harris Range* (named after its inventor) which was calculated by dividing the smallest number measured by the largest number measured. The Harris Range is related to the percent difference. If A is the smaller number and B is the larger number, we could define the percent difference as (B-A)/B. This simplifies to 1-A/B, and A/B is the Harris Range. The closer the Harris Range is to 1, the closer the two numbers. The groups in the section then used the Harris Range to compare the difference between the averages of two data sets or the variance within one set[1] (minimum trial/maximum trial).

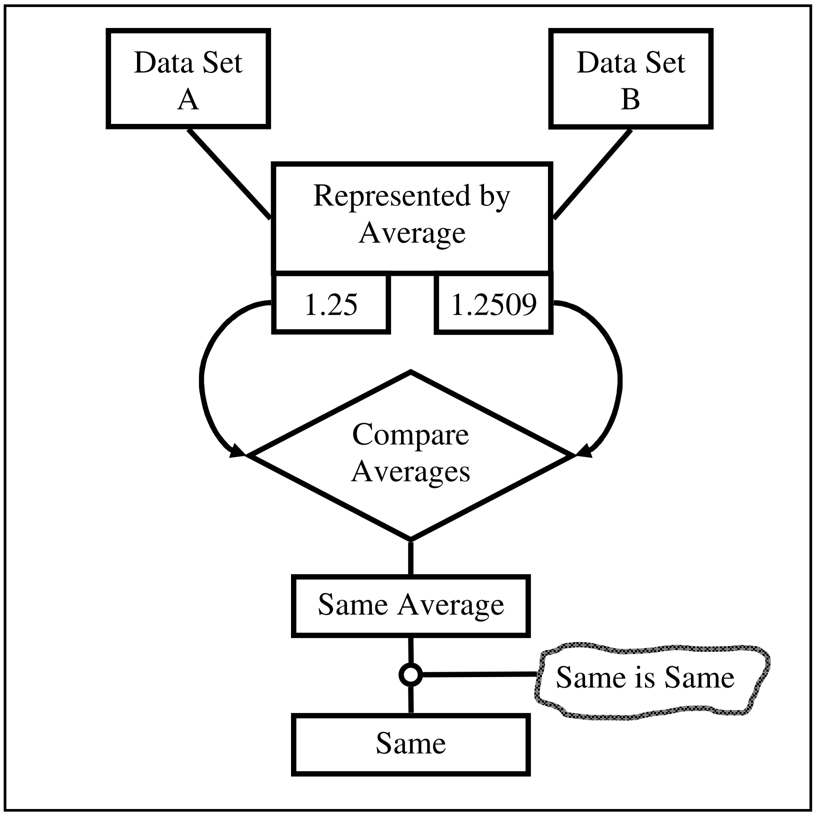


Figure 6-3

Same/Different argument representation: same average implies same cylinders

A slightly more complex argument is shown in Figure 6-4, where the students add in an idea which helps them make the decision that the two accelerations are different. We call this a controller because it controls the link between the comparison and the conclusion. This example comes from a group presentation during scientific community Lab 4 investigating whether two rings with different radii accelerate differently down a ramp.

---

[1] The section actually defined two different terms for this. They defined the Harris Range to compare the averages of two different sets of data, and the Judy Range to compare the maximum value and minimum value of one set of data. However, the students were inconsistent in how they used these terms, and often called them both the Harris Range, so that is what we do here.

> *1:* We have a 2 inch radius [ring] and a 2.5 inch radius [ring]. And we took some measurements, we just rolled it down the ramp, and we took the average [time] it took to get down the ramp, and then average velocity and average accelerations. We found that the acceleration of the smaller one was greater, it was a small difference I should say, 0.1 [m/s^2], but we felt that was notable because there was such a small difference in radius, it was only 2 and 2.5 inches, so we felt that it does affect it and if we had something like a 2 and a 6 inch radius we could see the difference much greater.

The student is claiming that the difference between the two averages is small, but it is significant: they do not expect a very large difference because the two objects are so similar. Other groups used an argument with a similar structure, but their controller focused on error:

> *2:* It could just be error, you know what I mean, like there's human error in the experiment, so like it could have just been a trial where someone didn't hit it quick enough, so like I think it's fine….I mean, since we're looking at the average, the average is close enough, because there is room for error when we took the average.

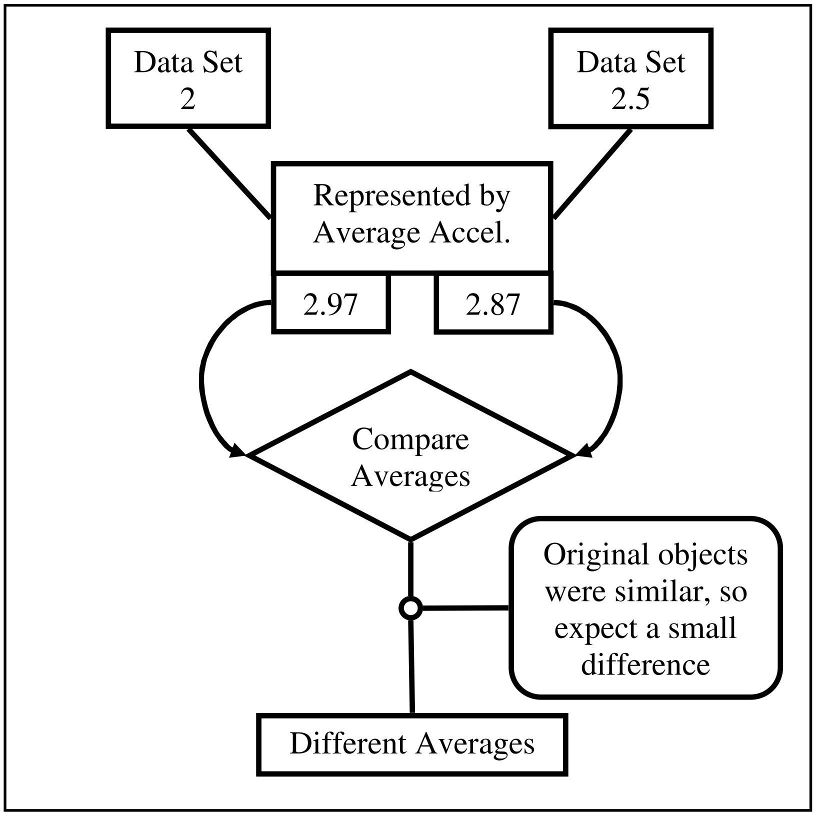


Figure 6-4

Same/Different argument representation with a controller

Instead of adding a controller, other students compared more than one aspect of the data set, such as average time, velocity, and acceleration. In this example the quantities are all related, but the students see this additional data as helping make a stronger conclusion. This creates more information for the students to compare all at the same horizontal level, producing a wider representation (see Figure 6-5). The group in the following example presented a comparison of three different cylinders, but we represent only cylinders 1 and 12 in the figure.

> *3:* We had the average time for each of them for 20 trials, we had 2.08, 2.03, and 2.07 seconds respectively, their average velocities at .48 m/s for these two and .49 m/s for number 12 were pretty similar. Their accelerations turned out to be .46 m/s^2 for 1 and 4 and .48 m/s^2 for number 12, so just based on looking at that we figured they're all the same.

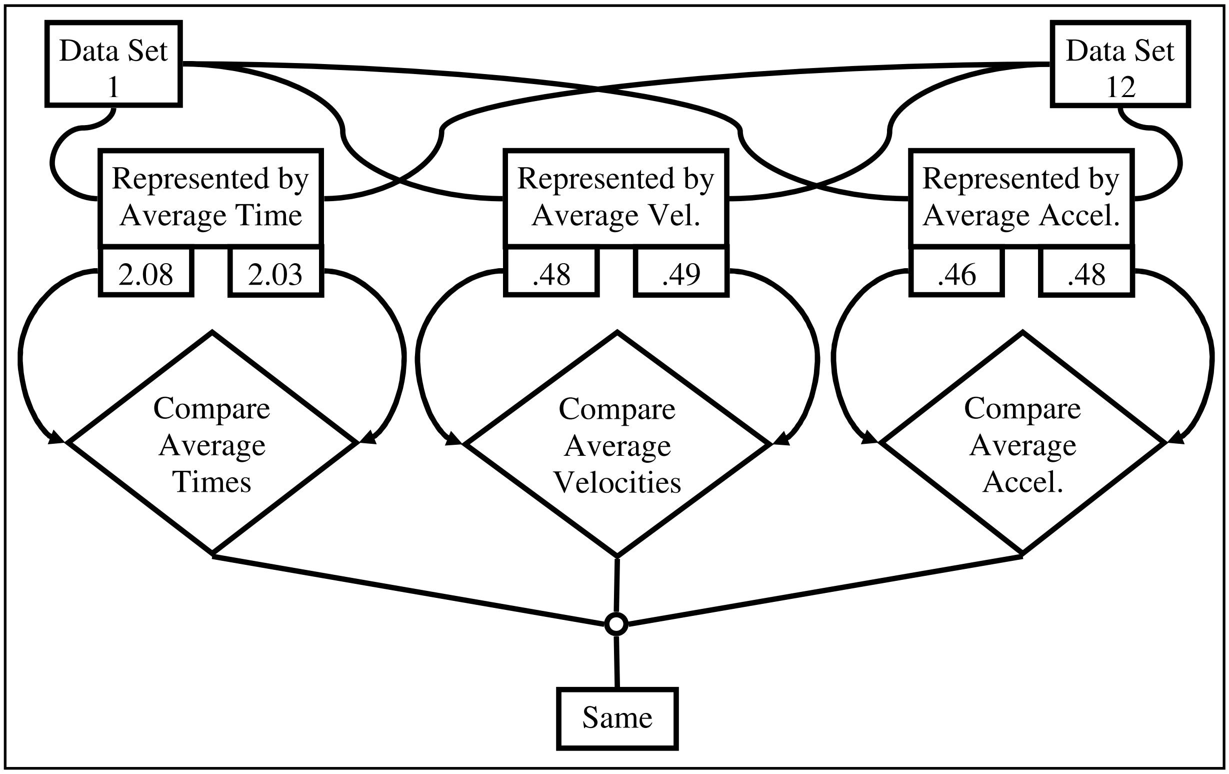


Figure 6-5

Same/Different argument representation: Using additional aspects of the data

Figure 6-3, Figure 6-4, and Figure 6-5 vary slightly in complexity, but all of them are primarily linear – each level leads to the next level only. In contrast, the most complex arguments use information from one level to inform two different levels. In the following example, students use the Harris Range (H.R.) to compare the variance within one data set (minimum/maximum) to the variance between the averages of two data sets (smaller average/larger average). They claim that if the data sets have a large range (if the H.R. within the data sets is far from 1) and if the averages are more similar (the H.R. between averages is closer to 1), then the averages really are the same.

1 *1:* The [closer to one] your Harris Range within your data, the more valid your data [averages] between the two [data sets] would be…

2 *2:* If [the Harris Range within the data set] was one, then the difference between [the two averages] would be the total, the difference between averages, you know? It's like

3 *3:* There would be no other

4 *2:* Exactly, there would be no other factors for this number, we could actually do this.
5 *1:* Right, but if it was a really large Harris Range within your data, and then your two final between trial datas [averages] are really far away also,
6 *4:* Maybe that would explain the difference.
7 *1:* You could use that to compensate, to say that well it could be the same because the data was so, it varied so widely within our data [sets].

In line 1, student 1 claims that a data set with a smaller range has a more valid average (the sophisticated idea of uncertainty in the mean). In line 4, student 2 explains this another way by saying if the data set had no range then the average would be exact, thus if the averages were at all different then the two data sets would be different – one could directly compare the two averages. In line 7, student 1 summarizes the argument that the difference within the data sets compensates for (or explains) the difference between the averages. The primary difference between this argument and the previous ones is the link from the data to the controller, as represented in Figure 6-6.

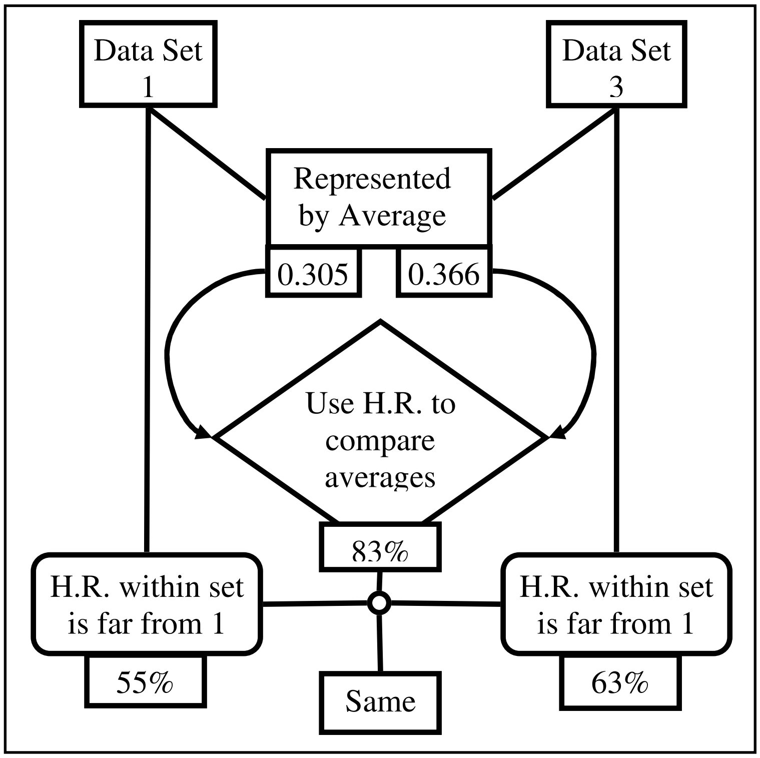


Figure 6-6

Same/Different argument representation: Controller based on data

Another example comes from a similar argument raised as a criticism of the group's presentation previously described in Figure 6-4. The group measured three trials of each ring, resulting in times of 4.60, 4.63, and 4.68 seconds for the large ring and 4.19, 4.62, and 4.84 seconds for the small ring. The student's criticism is that the difference between the averages should be larger than the difference within one data set to claim the averages are different:

> *4:* I'm saying just that to reach a conclusion that, shouldn't your difference between the two values for your average time at least be greater than the difference in your own ranges…I'm not completely convinced with what they've got because, again, going back to the values, for the smaller radius one of your values is 4.62 and for the bigger one one of your values is 4.63. Would you call that a good difference, or?

This student is making basically the same argument as that shown in Figure 6-6, except using absolute difference between trials instead of using the Harris Range to compare the numbers.

This representation has made it possible to analyze the flow of each argument and to compare the complexity of different arguments. We summarize those different complexities in Figure 6-7.

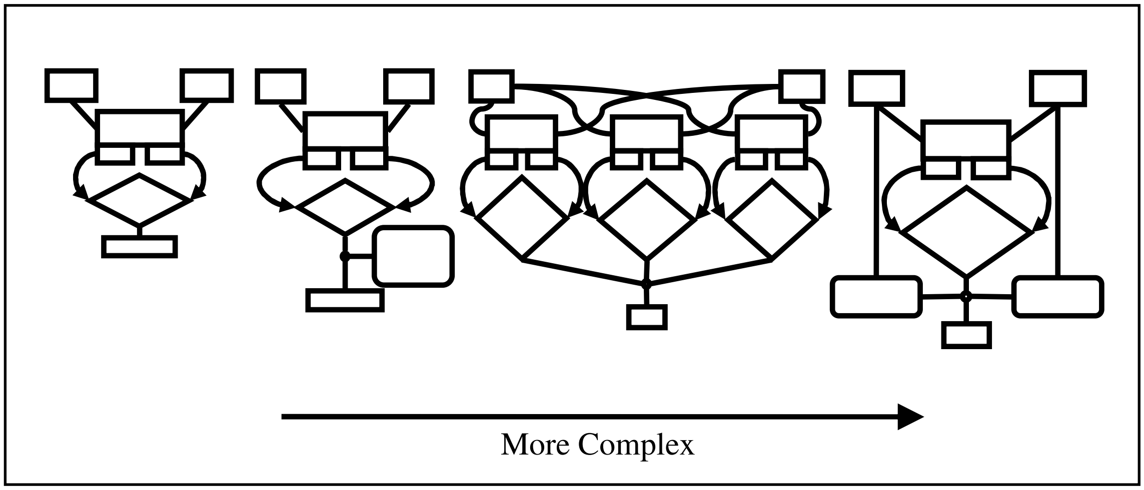


Figure 6-7

Different complexities of Same/Different arguments

A more complex argument is not necessarily a better argument: it depends on the context. There are situations where each of the arguments described here would be the most convincing arguments to make. For instance, if a professor is assigning final grades, she only uses the averages of students' grades, not their range. It would be useless for a student arguing for a higher grade to claim "well, at least I consistently scored poorly on the exams." However, if a student is deciding whom to cheat off of during an exam, he might take into consideration both the students' average exam score and their range of scores. It is not that we want students to stop making the simple arguments and to only make the complex arguments; we want students to be able to generate the appropriate argument and be able to evaluate all the different arguments.

Instructional Case Study

This representation is useful for the researcher and for the instructor: for studying student's arguments to make sense of what students are thinking and to see what intervention may be needed. The representation allows the instructor to parse students' arguments and determine if there are any pieces missing. In this section we study one group's argument as it evolves during scientific community lab 5, and discuss possible interventions.

This group of students investigated cylindroids 1 and 3 by timing 20 trials of each can rolling down a 122 cm long ramp. They calculated the acceleration for each trial, and the average acceleration and Harris Range for each cylindroid, (see Table 6-2).

	min trial	**max trial**	**average**	**min/max**
Cylindroid 3	0.266 m/s^2	0.482 m/s^2	0.366 m/s^2	0.55
Cylindroid 1	0.240 m/s^2	0.381 m/s^2	0.305 m/s^2	0.63

Table 6-2

The groups' data for lab 5

The Harris Range between the two cylinders was 0.83 (0.305/0.366), and by comparing the Harris Ranges within data sets to the Harris range between averages, the group concluded that the cylindroids were the same. See Figure 6-6 for a flow-chart of their initial argument, presented during class discussion. After the class discussion, the group wrote down their final decision (same), handed it to the TA, and then was told the answer. Before making their final decision, the group discussed what to do with the outlying data points.

1 *3:* I think we should do the [Harris] ranges without the outliers just to see.
2 *1:* I mean, even without that 0.482 [largest value obtained] it's still similar. It will be the same thing.
3 *4:* What's the highest without that?
4 *3:* It's 3.03 [seconds]. We could throw out all the three's, all the times, we'll be fine, like 3.03 [seconds, or 0.266 m/s^2], throw that out, and 3.19 [seconds, or 0.240 m/s^2], with one of them throw that out.
5 *1:* Well, even if we throw those out, it's just gonna make it more similar.
6 *3:* Exactly, it's to prove that our data is more similar, that's what I'm saying….just to make our data more convincing.

The group did not have time before handing in their conclusion to redo their calculations without the outliers, so they assumed (line 5) it would make the averages more similar. This was the first thing they tried after they found out that their answer was wrong, that cylindroids 1 and 3 were different.

7 *1:* What were the ranges?
8 *3:* Without the outliers, just to see, because we had just assumed-
9 *2:* Alright, without the outliers for 3 it's 0.323 and for 1 it's 0.273.
10 *4:* That's a big difference.
11 *3:* A huge difference.
12 *4:* See, that makes a big difference
13 *2:* Hoo, yeah. [laughs]
14 *3:* What are those now, these are the what?
15 *2:* The acceleration.
16 *3:* The averages between them?
17 *2:* [nods]
18 *4:* So we definitely needed to take the outliers…
19 *3:* We assumed previously before we should have actually did it.
20 *4:* So we have to say that [in our evaluation section]…They're so different, oh my god, I can't believe I didn't.

After throwing out the highest and lowest in each data set, student 2 added up the remaining 18 numbers and divided by 20 (a mistake) to get new averages of 0.323 and 0.273. Students 4, 3, and 2 look at those numbers and immediately claim the two cylinders are different (lines 10-12). However, if they apply their previous argument, the Harris Range between the new averages is .85, which is even closer than the previous .83. So the students must be doing something different to decide. Comparing the averages with the outliers (0.366 and 0.305) to the averages without the outliers (0.323 and 0.273) the main difference is that now the first digit is different. The difference between the averages has decreased (.061 to .05) and the Harris Range has increased (.83 to .85), both making a stronger argument that the cylinders are the same, yet the students claim these numbers prove that they are different. It seems plausible that the students are (tacitly) making the argument shown in Figure 6-8. The students never state that they are looking only at the first digit; because it is inferred, the box is dotted.

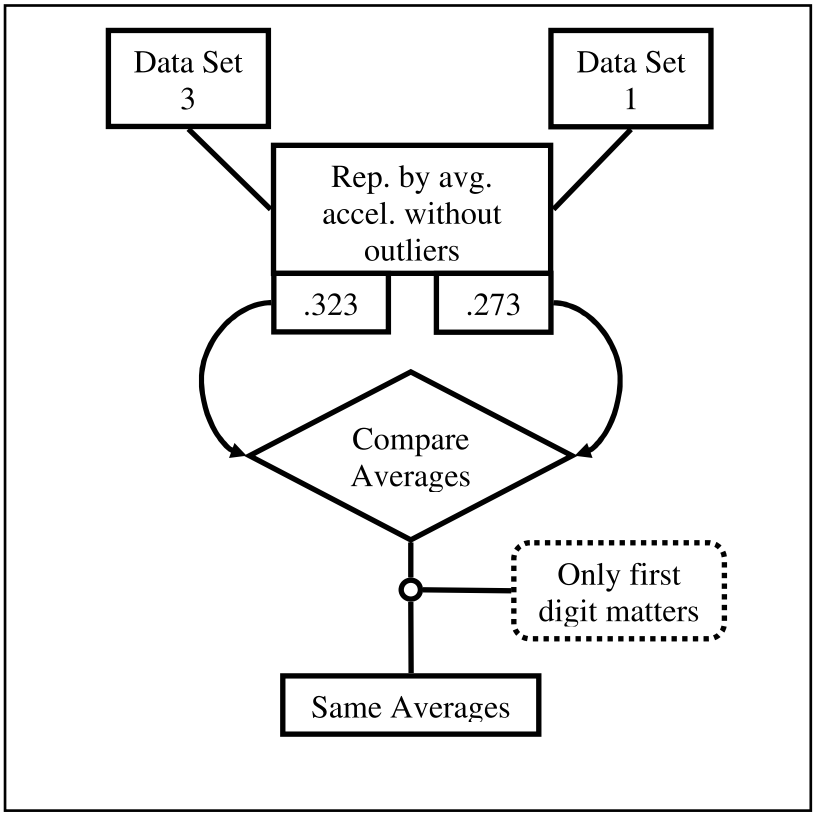


Figure 6-8

Different complexities of Same/Different arguments

There are many conceivable reasons for why the students switched to such a simple argument. On reason could be that this is the argument the students really believed all along and they used the other argument to make their presentation more convincing, as they state in line 6 and other times during the lab, and also during the lab quiz two weeks after this lab. Another reason could be that they are rushed for time: it is the last ten minutes of lab, they found out their previous reasoning was wrong, and this is the first thing that comes to mind that backs up the right answer.

Suppose the lab instructor noticed this change and decided to intervene. The first step might be to review the groups' initial argument and make explicit their second argument. If the group did not start to do this on their own, the instructor may want to encourage them to see if the two arguments are consistent. In this case the two arguments contradict each other, so she might encourage the students to refine one of the arguments. These two skills, seeking consistency and refining ideas, are useful for learning science in general, and are part of the larger curriculum surrounding the scientific community labs[1]. By having a clear idea of the students' arguments, the instructor is able to make decisions about what interventions might be beneficial.

Summary

In this chapter we use the method of knowledge analysis to study a few ideas underlying students' understanding of measurement and uncertainty. The data we analyzed was highly selected: first it had to contain discussion about measurement and uncertainty, and then it had to cluster with several other similar examples. This means we are making no claims about the frequency or lack of other ideas not found, only about the presence of the ideas we found. From analyzing videotapes of students working on the scientific community labs we found three common ideas: two resource-like, and one similar to a conception.

The *a constant has no effect* resource is used by students when controlling variables in an experiment, or in other parts of designing or evaluating an experimental method. Students may claim that the roughness of the table surface can be ignored, because both objects they investigated were rolled on the same surface. They may claim that background radiation can be ignored when calculating the half-life, because it's a constant value. Like all resources, this one is sometimes productive (it helps students control variables properly) and unproductive (students may ignore things that actually do have an effect).

Students use the *same is same* resource when comparing two things to decide if they are the same or different. They may study one or several properties of these objects, and determine similarity or difference from those properties: if the properties are the same, the objects are equivalent. This resource was used alone or used as part of a larger argument comparing two data sets. When the resource is used alone, students are often over-simplifying inappropriately and may need to use a more sophisticated argument.

Students' arguments about whether two sets of data are the same or different tended to be more conception-like, with an underlying structure. We use a flow-chart representation to detail the structure of student's arguments, and organize them according to complexity. In the most complex arguments students used one aspect of the data set (e.g. the range) to inform their decision on whether other aspects of the data (e.g. the averages) were the same or different. This representation is useful for instructors. It may help them to make sense of students' arguments, to determine whether the level of complexity is appropriate, and to decide what actions to take. As an example, we use the representation to describe how one group of students changed their argument during a scientific community lab, and to inform possible ways an instructor might intervene.

The flow-chart representation is also useful for studying students' arguments as a researcher: one could study how students' arguments change as a result of instruction or their evolution over time. One could also investigate whether a certain environment or lab context elicited more or less complex arguments from students. Unfortunately, such a study would be very difficult to perform on large numbers of students. In the next chapter we design a multiple-choice survey for studying students' ideas about gathering, analyzing, and comparing data.

---

[1] For more information on the larger curriculum, see the meta-learning course modifications in Chapter 3.

References

Buffler, A., Allie, S. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *Int. J. Sci. Educ., 23*(11), 1137-1156.

diSessa, A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction, 10*(2&3), 105-225.

Eimas, P. D., Siqueland, E.R., Jusczyk, P.W., and Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Sherin, B. L. (2001). How students understand physics equations. *Cognition and Instruction, 19*(4), 479.

Chapter 7: A Quantitative Study of Individuals' Understanding of Measurement and Uncertainty

Introduction

Interviewing individual students (Chapter 4) and analyzing videotapes of students working in the laboratory (Chapters 5 and 6) give useful information about a few students' interaction with the laboratory curriculum. To determine the effectiveness of the laboratory curriculum for all of the students requires a method that is easy to administer and analyze for large (over a hundred) numbers of students. In this chapter we describe the development of a multiple choice survey to test students' understanding of measurement and uncertainty. We then use this survey to assess the scientific community lab curriculum.

Our goal in this dissertation is to study how students build an understanding of measurement and uncertainty and how students use this to analyze their data. We have shown how the design of a laboratory can help this goal, by encouraging students to be in a frame where they use productive resources. We have analyzed how students use those resources to build an understanding of the underlying concepts. In this chapter we study how individual students use their understanding of the underlying concepts to analyze data.

The multiple choice survey is based on the Physics Measurement Questionnaire (PMQ), a survey developed in South Africa for testing students' ideas about collecting, processing, and comparing data. We first describe the development and analysis of this survey. We then describe the development of our multiple choice survey, which involved a cycle (performed twice) of refining and administering the questions and analyzing the results. The results of the final survey can be analyzed individually, with respect to students' responses to separate questions, or collectively, using the set and point paradigms. We describe and use both of these methods.

The Physics Measurement Questionnaire

The Physics Measurement Questionnaire (PMQ) is a free-response survey testing students' ideas about taking multiple measurements in distance and time, deciding on a final value to report, fitting data to a line, and comparing two sets of data. All the questions are based on the same experimental context, a ball rolling off of an elevated ramp onto the floor.[1] Students watch the demonstration of this context and then answer each question in order by choosing the response they most agree with and explaining their choice. Students are not allowed to change their answers to previous questions, so the last question of the survey allows them to note if they would like to change any of their previous answers.

Development

The PMQ was developed by Allie and Buffler at the University of Cape Town (UCT), South Africa and by Campbell and Lubben at the University of York, United Kingdom. It was designed for students entering the University of Cape Town who were identified (by their final high school

---

[1] See the section on Buffler and Allie in Chapter 2 for more information on the PMQ, including the experimental context, the survey questions, and the analysis.

exam) as underprepared in science. These students typically had poor previous science instruction with little or no laboratory experience and spoke English as a second language (Allie and Buffler, 1998). Because of this, the survey questions were designed to require as little reading as possible (see Figure 2-5).

Analysis

Students' responses are coded as coming from two paradigms: set or point (or mixed). In brief, the point paradigm is typified by the idea that measurement results in one value that could, in principle, have no uncertainty. When given five results from the same measurement, students reasoning from this paradigm may choose to report the first trial (because the researcher was most alert) the last trial (because the researcher had more practice) or the repeated value (if any measurements resulted in the same value). In contrast, the idea of the set paradigm is that measurement results in a cluster of values, each of which has some uncertainty.[1] Students reasoning from the set paradigm might choose to report the average of a set of data along with an uncertainty.

Preliminary University of Maryland Study

In the middle of the fall semester of 2001 we gave 92 students (all those taking the scientific community labs) a preliminary free response survey with questions taken from the PMQ. The purpose of this survey was to find out how Maryland students would answer and to determine which questions were useful for this new population. We suspected that several questions useful for Cape Town students would not provide useful information about Maryland students because all the students would answer the same way. Because of time constraints, we were only able to ask five of the probes: one about data collection, one about data processing, and three about data set comparison.[2] We used Buffler et al.'s (2001) analysis method, classifying each answer according to the choice the student made and whether their explanation indicated set, point, or mixed reasoning.

Maryland students more sophisticated

As expected, Maryland students' answers were more sophisticated than underprepared Cape Town students: Maryland students are primarily native English speakers and most have extensive lab experience from their freshman and sophomore college years.[3] For the first question about data collection, every student agreed with student 'C'. (See Figure 7-1.) Most students' reasoning was coded as set or mixed, with only two students coded as point. This is in contrast to the Cape Town students, where more than half of the students answered this question using point reasoning before instruction, and more than 20% answered a similar data collection question using point reasoning after instruction (Buffler et al., 2001). The two Maryland students whose reasoning was coded as point wanted to take measurements until the results became consistent. One example is as follows:

> "I would agree with student C on the bases that just a few trials may show exceptions rather than the rule. The trials should be conducted, if time is allowed, until they are getting a consistent d-value. This will allow them to better see the rule."

---

[1] See "New Measurement Concepts" in Chapter 3 for some examples of point and set reasoning.

[2] See Appendix F for the preliminary free response survey.

[3] For more on the student population, see Chapter 3.

The group of students decide to release the ball twice from $h = 400$ mm.

First release: $h = 400$ mm $d = 436$ mm
Second release: $h = 400$ mm $d = 426$ mm

The following discussion then takes place between the students:
A: We know enough. We don't need to repeat the measurement again.
B: We need to release the ball just one more time.
C: Three releases will not be enough. We should release the ball several more times.

Figure 7-1

The data collection probe

The second and third questions also had very few students using point reasoning (1% and 7%, respectively). The second question asked students to decide what to report for the final value of d. (See Figure 7-2.) The majority (98%) of students chose to report the average, and 33% of the students also included a range, usually the minimum value to the maximum value. Students taking the survey had already been through five scientific community labs focusing on the importance of range, so perhaps this encouraged them to report a range. Students' reasoning was primarily coded as set (98%) because they claimed that taking an average was the best way to represent the data.

The students continue to release the ball down the slope at a height $h = 400$ mm.

Release:	1	2	3	4	5
Distance (mm):	436	425	440	425	434

What should the students record as their final result for d?

Figure 7-2

The data processing probe

The third question asked students to compare two sets of data with the same average and different ranges, to determine which group got better results. Cape Town students were almost equally split when answering this question: 48% answered that the results are equally good since both groups got the same averages, and 53% answered that the group with the smaller range got better results (Allie et al., 1998). In contrast, the majority of Maryland students (92%) answered that the group with the smaller range got better results.

The last two questions (asking students to compare two sets of data) were the only ones which significant numbers of students answered using point reasoning. On each question, about 45% answered with set reasoning and about 45% answered with point reasoning, the remaining answers classified as mixed. The data sets in question 4 had a different average but the same size range, and in question 5 had a different averages and different size ranges that overlapped. A typical student response classified as set reasoning used ranges to make their conclusion, often drawing a number line to show the overlap.

> "Even though their averages are different, their ranges almost completely overlap. Because the ranges are so similar, I believe that the results of both groups' data agree."

A typical point response compared the two averages and judged whether they agreed or disagreed based on the students' experience and not on the data.

> "I think they agree only because the different [between the two averages] is two millimeters (which is very small). It's so close that I don't think the difference is significant. Maybe the way they measure was a little different."

When we combine the results of the last two questions, about a third of the students answered set on both, a third answered point on both, and a third answered using mixed reasoning.

We combine students' answers to all five questions by assigning 10 points for a set answer, -10 points for a point answer, and 0 points for a mixed answer. Thus, students who answered all questions using point reasoning would receive a score of -50 points. The highest number of students, 16, answered all the questions with set reasoning (see Figure 7-3), and no students answered all the questions with point reasoning. Because the results are saturated, we refined our choice of questions for the survey we gave the following semester.

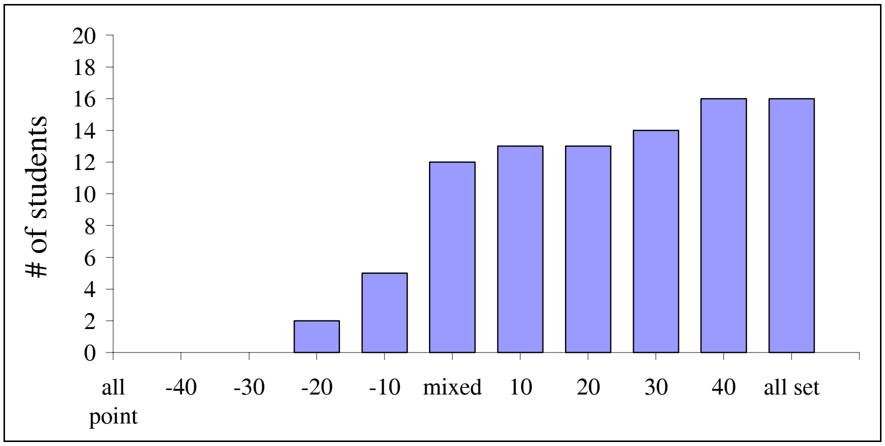


Figure 7-3

The number of students answering with set or point on the preliminary survey

Free-Response Survey

In the spring of 2002 we gave a survey to students in the electricity and magnetism (second semester algebra-based physics) laboratory before and after the semester. The six questions were again taken from the PMQ, but were chosen to differentiate more among students than those in the preliminary survey. We also added two questions adapted from Sere et al. (1993) about combining the results from two groups' measurements.[1] We use the results from this survey mainly to develop the final version of the survey, the multiple choice survey, and to design a new analysis method.

At the University of Maryland, the first semester introductory physics course is divorced from the second semester course, so students switch professors between semesters according to their schedule or personal preference. Out of the 79 students who took both the pre and post survey, 43 were in the scientific community labs in the previous semester (we call these the "old" students) and 36 were in traditional labs in the previous semester (we call these the "new" students). This means that the pre-survey acts like a post-survey for the old students and a pre-survey for the new students. However, students are self-selected, so the results must be somewhat carefully interpreted. This reason, in addition to the relatively small numbers, makes it difficult to draw any

---

[1] The free-response survey is included in Appendix F.

broad general conclusions from this survey's data. We use students' responses on this survey primarily to help develop the answer choices for the multiple-choice survey.

Analysis

The free-response survey was coded slightly differently than the preliminary survey. Instead of coding each response as set, point, or mixed and then classifying it further according to the specific reason, we reversed those two steps. Students' answers were relatively consistent and could be classified as belonging to six or fewer categories. Responses were first classified as belonging to one or more of those categories, and then the categories were identified as set, point, or mixed. For example, questions 3, 4 and 5 on the survey asked students to compare two data sets given the average and either the individual data points (question 3) or the standard deviation (questions 4 and 5). The categories for each question were the same, and all of the student responses fit into one or more of the categories. The categories, in order of most frequent to least frequent, are described in Table 7-1. Students' responses were sometimes coded in multiple categories. For example, the following student response was coded both as *Compare Averages* (because they compare the two averages, 434 and 442) and *Range Size* (because they compare the standard deviations):

> "The results do not agree because the standard deviations are different and the numbers (434 and 442) are significantly different."

Name	Short Description	Example
Range Overlap	The two data ranges do/do not overlap.	I would say the results do agree, because the ranges overlap a fair amount.
Compare Averages	The averages are the same/different.	The results are different because group B's distance is larger than group A's.
Average in Range	The average of one set is/is not in the range of the other set.	Because the average of the two results do not fall within each other's standard deviation, I would not say they agree.
Value Comparison	The values from each set are the same/different.	Group A has 3 scores that are identical to group B. The average may be different, but the results are the same.
Range Size	The two sets have the same/different size range.	Their range is similar [for two non-overlapping sets with same standard deviation].

Table 7-1

Coding categories for questions 4 and 5 on the free-response survey

Results

Using the coding scheme shown in Table 7-1 we compared students' answers from pre to post and between old students and new students. For the old students, no answer category increased or decreased by more than 4 students (9%) from pre to post. Also for the new students most categories remained very similar from pre to post (at most changing by 8 students). The only significant difference occurred on the question asking what value to report as the final result. In the pre-survey, 4 new students (11%) and 17 old students (40%) reported a range as well as an average. This increased to 14 new students (39%) and 19 old students (44%) reporting a range in the post-survey (see Figure 7-4). Comparing the old and new students in the pre-survey, and the pre and post new students, it may be that students are becoming more likely to report a range. The number of old students reporting a range did not change significantly from pre to post.

Of the categories in Table 7-1 two are identified as set (*Range Overlap* and *Average in Range*) because these answers indicate that the students are considering the spread of values from the measurement. The other categories were coded as point (*Compare Averages, Value Comparison*, and *Range Size*) because the students are treating each number (the average, data points, or the standard deviation) as a single point. However, here is where the problem with set and point coding rears its head.[1] Concluding that two data sets agree because they have three identical values is not necessarily wrong, even though it is coded as set: such data sets would be likely to have overlapping ranges. Also, there are times when it is important to consider the size of the standard deviation when comparing two data sets.[2] It is not true that point reasoning is always incorrect and set reasoning is always correct.

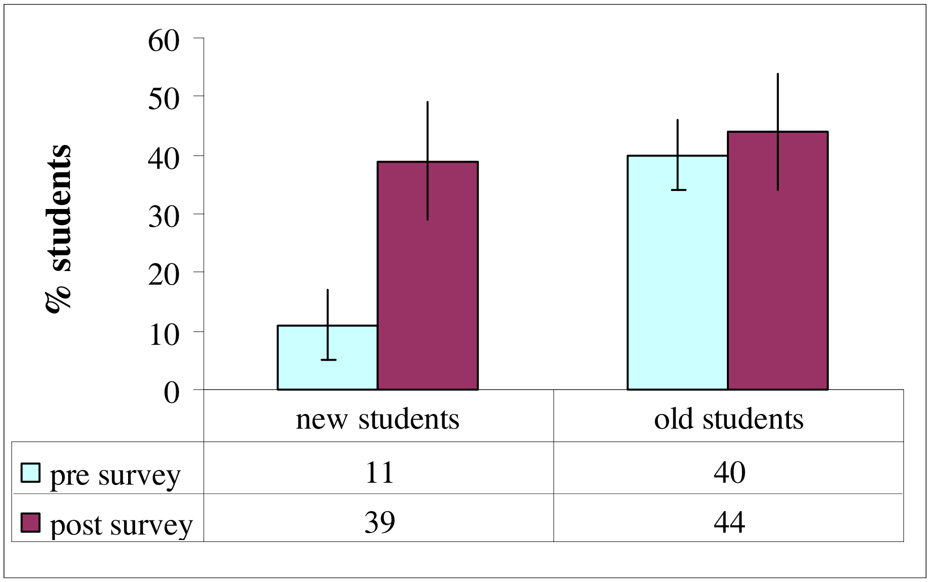


Figure 7-4

The number of old and new students reporting both an average and a range.

By assigning positive values to each set answer and negative values to each point answer, as in the preliminary survey, we were able to determine how many students were reasoning from mostly point, mostly set, or mixed paradigms. Looking at the pre-survey (Figure 7-5) it appears that the questions still may not be differentiating well between students because the results are saturated at the set end. If we compare the pre-survey (Figure 7-5) to the post-survey (Figure 7-6) it seems that more students are using more set reasoning. But we do not necessarily want students to move to all set reasoning and stop using any point reasoning.

To distinguish between students who increase in set reasoning from students who decrease in point reasoning, we introduce a scatter plot representation. Both set and point answers are assigned positive values, and each student is represented by a dot with two coordinates (set, point). (See Figure 7-7.) By comparing pre and post responses, we can determine whether students are changing their set reasoning separately from their point reasoning. Looking at Figure 7-7 and Figure 7-8, it appears that students are increasing both their set reasoning and their point reasoning.

---

[1] See chapter 2 for more on the problem with set and point.

[2] For example, if testing the contents of an unlabelled can by rolling it down a ramp, a can of beans will have a much larger standard deviation than a can of tomato paste, but the averages for each can will not be significantly different.

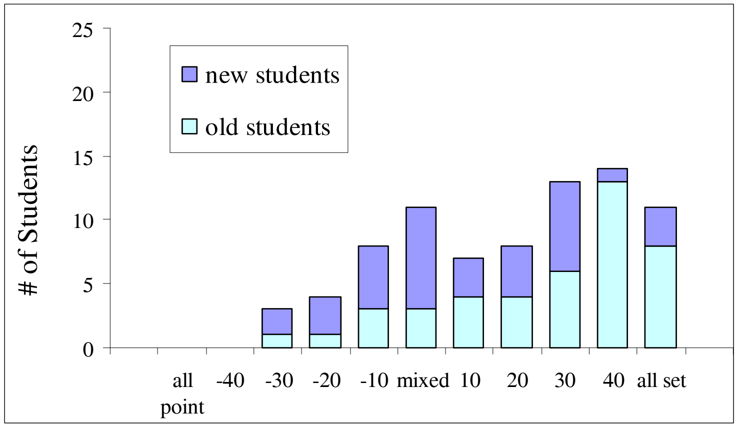


Figure 7-5

Number of students using set or point on the pre-survey

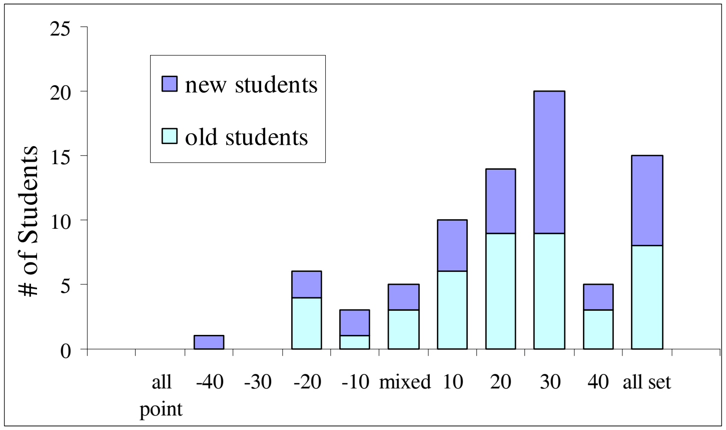


Figure 7-6

Number of students using set or point on the post-survey

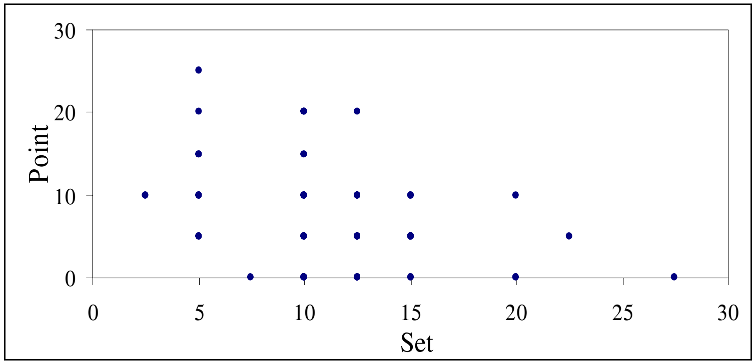


Figure 7-7

The scatter plot for the free response pre-survey

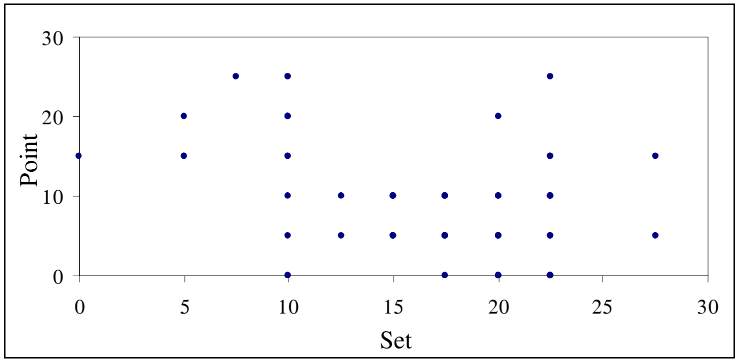


Figure 7-8

The scatter plot for the free response post-survey

Multiple-Choice Survey

The time it takes to administer, code, and analyze a free-response survey is prohibitive, so we used the responses and categories from the free-response survey to develop a multiple-choice survey. A multiple-choice survey allows one to study the impact of a certain instructional technique on large numbers of students in a relatively easy manner. This format also makes dissemination of the survey possible, because of the relative ease of administration and analysis. In this section we describe the development of this survey and then the results from the Fall semester 2002 scientific community labs.

Development

The questions on the multiple-choice survey are primarily the same as the free-response survey. The survey has one question on data processing, five questions on data comparison (three using data points and two using standard deviation), and two questions on combining data.[1] The answer choices were mostly taken from the categories developed for the free-response survey. We wrote a typical answer for each category and ordered them from least complicated to most complicated. For example, in the question in Figure 7-9, choices 3 and 8 come from the *Compare Averages* category, choices 4 and 10 come from the *Range Overlap* category, and choice 9 comes from the *Range Size* category.

There are certain problems with a multiple choice survey[2]. One of them is that students may just choose the response that sounds most "scientific" instead of the response with which they most agree. Another problem is that a student's answer may not be one of the choices. We attempt to reduce both of these problems by administering the survey electronically. The student views the question and chooses an answer: for the question in Figure 7-9, a student had to first choose 'yes' or 'no'. She then would be shown a window of typical reasons for that answer and asked to choose one or more reasons or to type in her own reason. This forces students to think about the question and decide on an answer before seeing the choices. It also solves one problem of the paper and pencil survey, that students were able to go back and change answers after working through more advanced questions.

Results

In the fall semester of 2002 we gave students the multiple choice survey before and after taking the scientific community labs. Students received participation points for filling the survey out online and 120 students submitted both the pre and post survey out of about 160 students enrolled in the class. In this section we start by analyzing the responses to each question, and then look at overall changes.

Improvements on separate questions

We begin by discussing specific answer choices which changed by 10 or more students from pre to post. For those interested, all of the results can be found in Appendix F. The number of students who chose to answer 'other' was low, with an average of two students and at most six for any question. This may indicate that most of students' answers were included among the available choices, though it could be that the students saw at least one of the answers as a plausible (if not the best) reason and declined to type in their other, primary reason.

---

[1] The multiple choice survey is included in Appendix F.

[2] See the discussion on surveys in Redish (2003) for more information.

Two other groups of students compare their results for d obtained by releasing the ball at $h = 400$ mm. Their means and the standard deviation of the means for their releases are shown below.

Group A: $d = 434 \pm 5$ mm
Group B: $d = 442 \pm 6$ mm

Do the results of the two groups agree?

Yes

1. There isn't a significant difference between the two group's results.
2. Everything has error, it's impossible to get exactly the same every time.
3. There's a difference of eight millimeters between the two group's averages. Eight millimeters is small, and so they pretty much have the same result.
4. Group A's range is from 429 to 439, group B's from 436 to 448, so the ranges overlap.
5. Other

No

6. Their averages are different.
7. There is a significant difference between the two group's results.
8. The difference of eight millimeters is a large difference compared to the distances they're measuring.
9. Group A's range has a width of 10 mm, group B's range has a width of 12 mm, so they have different results.
10. Group A's range is from 429 to 439, group B's range is from 436 to 448, they only overlap for 3 mm which is not enough.
11. Group A's average of 434 does not fall within the range for group B (436 to 448), and vice versa.
12. Other

Figure 7-9

A data comparison question from the multiple-choice survey

The first question presented students with five data points and asked them to write down what to report as the final result (Figure 7-1). As expected, more than 75% of the students gave the average. The number of students reporting the range increased from 2 to 19 pre to post. Unfortunately, this is still only 16% of the students reporting a range, but it did increase. An interesting, though perhaps insignificant result, is that the number of students who included units in their answer decreased from 15 to 9.

The largest changes occurred for the choices in the *Range Overlap* category for the data comparison questions. The number of students selecting these choices increased from 14 to 52, 19 to 53, and 39 to 69 for questions 3, 4, and 7 respectively. Comparing sets of data by range overlap was a recurring theme in almost all of the scientific community labs and the primary goal[1] of labs 2, 3, and 4. Students seem to at least be able to recognize that this is the answer to choose, sometimes to the extent of refusing more scientific sounding choices such as those including standard

---

[1] See Chapter 3 for more on the goals of each lab.

deviation. Question 6 asked students to combine two groups' results given the average and standard deviation. The number of students choosing "they should put all their data points together and use those to calculate an average and a standard deviation" decreased from 58 to 24 and the number of students choosing "they should report the average with the range of both group's data" increased from 21 to 42. It may be that students were discouraged from using standard deviation. In the scientific community labs, whenever students used a new tool they had to define and explain the tool so that the other students understood. Most students who used standard deviation had it built into their calculator and could not explain it, so only one section out of eight had standard deviation defined on their tools and terms board.

The number of students selecting the lowest level choices for the data comparison questions ("there is/isn't a significant difference between the two groups' results", "everything has error, it's impossible to get exactly the same every time", and "their averages are different") reduced from pre to post for every question, often significantly. Students were allowed to select multiple answers because in the free response survey students often wrote down multiple reasons. They could have selected one of these choices as well as other answers, but it seems that many students no longer agreed with these choices in the post survey. The lab TA often responded negatively to these types of answers when they were given during the class discussion or written in the lab report, by stating that those answers are not convincing, or they require more information such as a way to define significance. The survey results seem to indicate that this message was communicated to the students.

Overall Results

In addition to looking at the changes in specific question choices, we also analyzed the results from the survey in terms of set and point in a similar way to the free-response survey discussed previously. Each answer choice was designated as set or point and assigned a value from 1 to 3 depending on how strongly they were set or point. For example, range overlap answer choices were assigned as set with a value of 3, and compare average choices were assigned as point with a value of 3. The answer choice "There is/isn't a significant difference between the two group's results" was assigned a value of one, and set or point depending on the question and whether it was likely that students were comparing averages or comparing range overlap. The values for each answer choice are included in Appendix F.

For the histogram representation, we sum the value of a student's answers counting set answers as positive and point answers as negative, and we report the sum of every student's answers. The results from the pre-survey (Figure 7-10) show a full distribution – no longer are the results saturated, at least on the pre-survey. The post-survey (Figure 7-11) does seem to show some saturation. From pre to post students are shifted to the right (Figure 7-10 and Figure 7-11) so students are answering using more set reasoning or using less point reasoning.

For the scatter plot, we added students' set answers separately from their point answers and assigned each student two coordinates (total set, total point). Both the pre-survey (Figure 7-12) and the post survey (Figure 7-13) show a slight negative correlation between set and point, in contrast to the free-response survey results. Thus students who answer a lot of questions using set reasoning are not answering many questions with point reasoning. On Figure 7-13 the average of all the points for the pre survey and the post survey are shown as triangles. From pre to post in general the amount of set reasoning increased but the amount of point reasoning decreased by only a small amount. This change is actually appropriate: we want student's set reasoning to increase, but we do not want them to completely drop their point reasoning. Students should be able to use set reasoning or point reasoning, deciding which would be most fruitful.

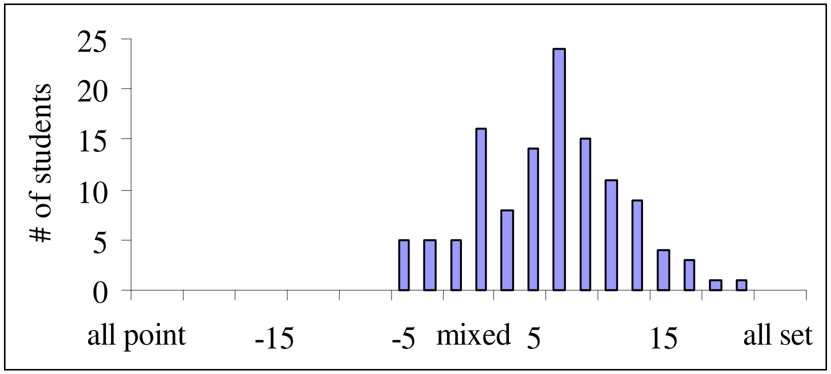


Figure 7-10

Number of students using set or point on the mult. choice pre-survey

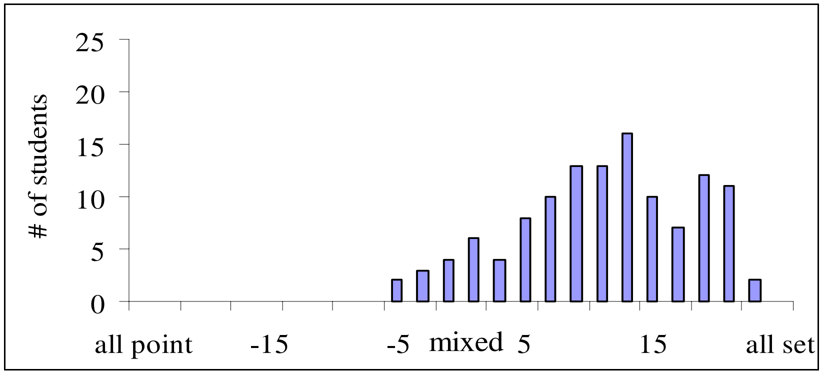


Figure 7-11

Number of students using set or point on the mult. choice post-survey

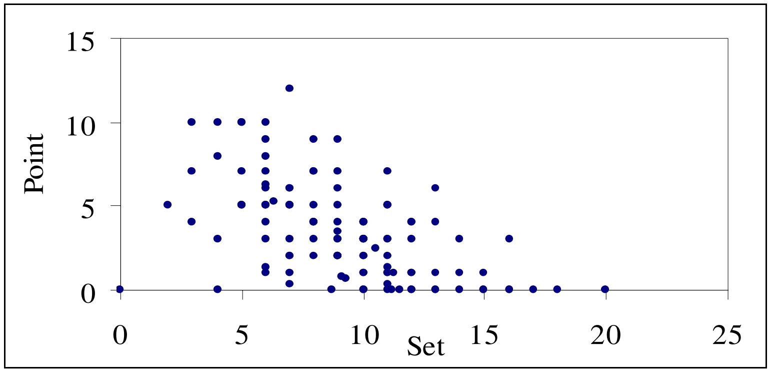


Figure 7-12

The scatter plot for the multiple choice pre-survey

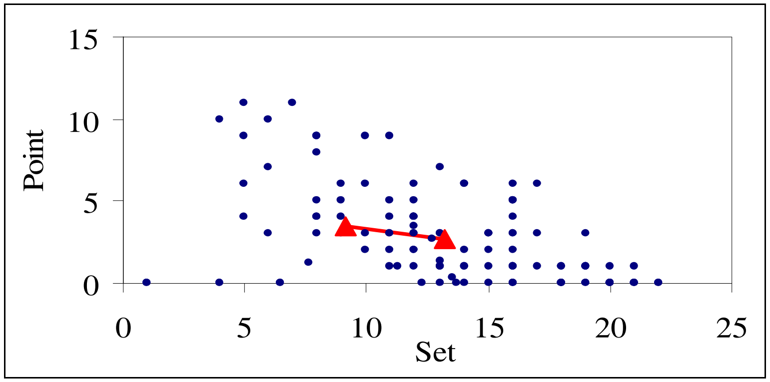


Figure 7-13

The scatter plot for the multiple choice post-survey

Summary

In this chapter we describe the development of a multiple-choice survey for testing students' ideas about gathering, analyzing, and comparing data. We then use this survey to assess the effectiveness of the scientific community labs in teaching students these skills. We based the survey on the Physics Measurement Questionnaire, a free-response survey developed by Buffler et al. (2001). Students at the University of Maryland were more sophisticated than the original PMQ population, so we eliminated some questions which did not differentiate among students and added two questions on combining data sets to create a free-response survey.

The results from the free response survey were analyzed by first sorting them into categories and then coding them as set or point. These answer categories informed the creation of an eight question multiple choice survey, defining the different answer choices. We designed the multiple choice survey to be administered electronically so that students cannot go back and change their answers to previous questions and so that students are shown the reasons only after they have decided upon an answer (such as "the two data sets agree"). The multiple choice survey can be analyzed in two ways: by looking at the overall changes in student's set and point reasoning, and by looking at changes in responses to individual questions.

The multiple-choice survey was administered to 120 students in fall 2002 before and after the scientific community labs. Both the results from the individual questions and the overall results show that students' set reasoning seems to be increasing after taking the scientific community labs. Students are less likely to give lower level incomplete answers in the post survey, perhaps indicating that they are selecting only "convincing" answers, as encouraged in lab. Students are also more likely to use range overlap when comparing data, which also was stressed in lab.

The overall results show that students' point reasoning decreased slightly and set reasoning increased. One goal of the scientific community labs was to teach students to use either set reasoning or point reasoning as appropriate. The results from the survey show that one part of this goal, for students to learn to use set reasoning, seems to have been accomplished.

References

Allie, S., A. (1998). A course in tools and procedures for Physics 1. *Am. J. Phys., 66*(7), 613-624.

Allie, S., Buffler, A., Kaunda, L., Campbell, B., and Lubben, F. (1998). First-year physics students' perceptions of the quality of experimental measures. *Int. J. Sci. Educ., 20*(4), 447-459.

Buffler, A., Allie, S., Lubben, F., and Campbell, B. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *Int. J. Sci. Educ., 23*(11), 1137-1156.

Redish, E. F. (2003). *Teaching Physics with the Physics Suite*. John Wiley & Sons, Inc.

Sere, M., Journeaux, R., Larcher, C. (1993). Learning the statistical analysis of measurement errors. *Int. J. Sci. Educ., 15*(4), 427-438.

Chapter 8: Implications and Applications

Introduction

This dissertation was written with several audiences in mind, including someone teaching a physics laboratory, someone working to design a laboratory curriculum and someone studying students' reasoning in the laboratory. The laboratory teacher is likely to be more interested in ways to improve their teaching, the laboratory designer in ways to evaluate and improve the curriculum, and the researcher in tools and methods for studying students' reasoning. In this chapter we summarize particular elements of the dissertation which are of interest to specific audiences, giving suggestions tailored for that audience. We first address the interests of the laboratory instructor, discussing how to fully implement the scientific community lab curriculum. We also give suggestions for improving traditional labs for those who cannot completely change the curriculum. For the laboratory designer we explain several measurement concepts, often ignored in laboratory curricula, which we have shown to be critical for understanding uncertainty. We then review the research tools used in this dissertation and discuss ways to adapt them for evaluating other laboratory curricula (for the laboratory designer) and for studying other student populations and other research questions (for the researcher).

Suggestions for fully implementing curriculum

In Chapter 2 and throughout this dissertation we argue that understanding measurement and uncertainty is critical for students preparing for careers in science, and useful for everyday life. In the traditional laboratory, students may see uncertainty analysis as useless busy work – when asked what they would change about the lab, more students suggested removing all the uncertainty analysis than suggested relating the lab to lecture. For students to see the purpose of uncertainty analysis and for them to build an understanding of the concepts necessary requires a completely restructured laboratory design. The scientific community labs are one attempt at this.

Each scientific community lab begins with a brief class discussion designed to elicit the main measurement concept for that lab. Students work in groups of four to design and implement a method for answering a laboratory question such as "Does friction depend on surface area?"[1] In the last half an hour of the two hour period, each group presents their method and conclusion while the rest of the students ask questions and critique the presented design. During this discussion students may propose a new tool or term they found useful. If accepted by the class, this is then written on that section's *tools and terms board*. Each week the students hand in a group lab report, called a *weekly log*, describing and evaluating the steps they took toward solving the problem. Twice a semester students individually take a one hour *laboratory quiz*[2] containing short answer questions on the measurement concepts and tools and terms and an essay question asking them to evaluate a lab design.[3]

We have implemented the scientific community labs in two different semesters at the University of Maryland and evaluated the curriculum using surveys, interviews, and video analysis. The results of an anonymous survey show that more students in the scientific community labs self-report that they learned skills such as problem solving, critical thinking, and experimental design and that these skills would be

---

[1] The lab handouts for the students are included in Appendix A and for the TA's are in Appendix B.

[2] A sample lab quiz is included in Appendix C.

[3] See Chapter 3 for a detailed description of the scientific community lab curriculum, including a description of the measurement concepts and the reasons for each particular curriculum element.

helpful to them in the future (66% versus 11% in the traditional labs).[1] An analysis of students' behavior during lab showed that students in the scientific community labs spend more time sense-making (21% of the time versus 4% in traditional labs).[2] The number of students who compared data sets by looking at how the ranges of the data sets overlapped increased from before to after the scientific community labs, doubling or more on several questions.[3] These and other results seem to indicate that the scientific community labs are making a positive impact on the students in the algebra-based introductory physics course at the University of Maryland. In the rest of this section, we give specific warnings and suggestions about implementing the scientific community lab curriculum elsewhere: scheduling the labs, the importance of the homework questions, and the teaching load.

Scheduling

The scientific community labs were designed for twelve 110 minute periods. During most periods the labs were very rushed, and the final class discussion often had to be cut short. A two and a half hour lab period or a three hour lab period would likely work better, and would require no changes to the labs. Any shorter time would require two periods to complete a lab. It is also important that the labs are completed in order, because each lab builds on the skills and concepts of the previous one. (See Figure 3-5 in Chapter 3.)

Students should work in groups of three or four, a group larger than the pair of students or single students often working in traditional physics labs. One reason for this is that we are asking students to do a difficult task which requires a lot of creativity to design a method, hands to carry out the method, and thinking skills to evaluate the method. Another reason is that the scientific community labs are designed to be a social situation where students argue with and convince other students. The large group size encourages students to interact. The third reason is more logistical: every group must have time to present in the final class discussion and a smaller group size would require more presentations which would take more time.

Two of the periods are used to administer a laboratory quiz[4], a crucial component of the labs. Students take the lab quiz individually, which encourages them to build their own understanding during the lab instead of relying only on their group members. It also communicates to the students that the laboratory is teaching serious topics which they are expected to understand – the laboratory is not just a requirement which they must attend. The lab quiz gives students individual feedback on their reasoning, and the solutions provide examples of sophisticated reasoning. Anecdotally, TA's have noticed a significant improvement in students' reasoning during class and in written lab reports after the lab quizzes.

Homework versus Pre-Labs

Traditional labs at the University of Maryland give students 2 points out of 20 on their lab report grade for completing a set of questions before lab, called the pre-lab. The TA walks around during lab and checks to see which students have completed the pre-lab, giving them two points for completion. The first version of the scientific community labs similarly used a pre-lab to introduce a measurement concept. This did not work very well, for the following reasons:

- Checking the pre-lab during lab required a lot of TA time needed for other things and interrupted students' discussions, often leading them off-task.

---

[1] See Chapter 4 for more results from the anonymous survey.

[2] See Chapter 5 for more information about the students' behavior during lab.

[3] See Chapter 7 for the development and results of the measurement survey.

[4] A sample laboratory quiz is included in Appendix C.

- Students viewed pre-labs as busy work to be quickly completed when time allowed, often right before lab or during lecture.
- Students complained about having too many separate assignments to remember for one physics course.

Writing the pre-lab as one homework question included in the weekly problem set assigned during lecture solved these problems. It may be difficult to coordinate the laboratory and the lecture, but it is important. Many of the lab homework questions[1] make use of physics concepts needed in the lab, giving students practice in both physics and measurement concepts. Several of them ask about measurement concepts only, and most of them are essay questions. We recommend that they be graded on effort only and not correctness, because they are asked before the concept is discussed in lab. (There are also lab homework questions for use after lab, and those can be graded for correctness.)

The lab homework introduces a measurement concept, and is often used during the class discussion at the beginning of the lab. It does not force the students to read the lab handout, as a traditional pre-lab does. After the first lab, students recognize that they must do this on their own, and with a little encouragement, often do. This is shown by a transcript excerpt from the second scientific community lab:

1: Yeah, next time I'm really gonna sit down and think about how we're gonna do it.
2: I know, so I know what to do: "wait, I have the plan!"

This is one example where the proper design of the lab causes students to do what we wish without forcing it upon them by making it an explicit part of their grade.

Teaching Load

Even those who agree with the goals of the scientific community labs may be hesitant to implement them because of a lack of funds or cost in person-hours. In this section we compare the time it takes a graduate teaching assistant to run a traditional lab to the time it takes to run the scientific community labs. At Maryland a typical full time teaching assistant works 20 hours a week and is responsible for three introductory laboratory sections. A TA in a traditional lab has many of the same time requirements as a TA in the scientific community labs (see Table 8-1), except for grading the laboratory reports. Because students in the scientific community labs hand in group weekly reports, there are one quarter as many reports to grade (assuming groups of four). This allows TAs to spend more time grading each report and less time grading overall. Another primary difference is in TA training, which requires more time because of the extra training needed to learn how to interact with groups and run a whole-class discussion.

Lab	Contact Hours	Training	Grading Reports	Office Hours	Lab Quiz Grading	Exam Grading
Traditional	6	1	10	2	0	1
Sci. Comm.	6	3	6	3	1	1

Table 8-1

The weekly workload for a laboratory teaching assistant

The total teaching load for the TA is the same, but the load on the professor or head TA is increased because they must prepare for and run the extra training. However, the equipment set-up for the scientific community labs is less expensive and time consuming than the traditional labs. Because the students

---

[1] All of the lab homework is included in Appendix C.

design their own experiment, most equipment is available all the time, such as meter sticks, stop watches, assorted weights, string, etc. Traditional labs, because they must allow students to collect data which will prove a theory, typically have expensive equipment developed specially for one lab. It takes less time to set up equipment for a scientific community lab, but more time to prepare the instructors. Taking everything into account, the time load for the scientific community lab is similar to that for a traditional lab.

Suggestions for traditional labs

As discussed in Chapter 4 and elsewhere, the traditional cookbook laboratory often causes students to be in an unproductive frame, where they are more concerned with following the procedure and completing the lab than with understanding concepts, whether physics concepts or measurement concepts. Attempting to change this frame by completely redesigning the lecture, discussion section, and laboratory was mostly successful. However, if one is unable to make such drastic changes, there are still a few changes which may improve the traditional laboratory. We discuss these changes in this section.

Remove Theory Validation Questions

One detrimental part of the traditional laboratory is that students see themselves as seeking "the right result". If there is a right answer and students must get it, there is no reason for students to interpret their results or try to make a convincing argument. Students may view the pedagogical goal of lab as proving to them that the physics theory works rather than understanding the physics concepts. These students may then conclude that lab is useless, as shown by the words of a student: "The teacher tells me this is how it is, I accept it. I don't need to go to lab and prove it."[1] In traditional labs it may be possible to change the lab slightly, using the same method and the same equipment, but asking a different question with a more open answer.

One traditional lab on waves and vibrations in the introductory physics course at the University of Maryland asks students to find the frequency of a reed vibrating a string by varying the tension in the string and measuring the nodes. This question could be changed and given a real life context as follows: "You are on a diet and need to measure out certain amounts of food to eat using a scale. If you use the device and method described below, how well can you measure a food's weight?" This type of question gives students the method for using the apparatus but does not give them the method for determining their uncertainty, leaving students with something to think and argue about. Many other cookbook lab questions can be changed in this manner.

Ask "Explain" Questions

Unfortunately, students in the traditional lab frequently do not monitor how well they understand, or are in a frame where they do not expect to understand. Asking students to explain what they are doing may help increase students understanding. Schoenfeld (1987) uses three questions for this:

- What are you doing?
- Why are you doing it?
- How does it help you?

As discussed in Chapter 2, the student's frame is very important. Asking students to do this monitoring when they are not in a frame where they view monitoring as acceptable behavior is likely to fail. Just telling students "in this lab we want you to think and make sense of things" without designing every aspect of the lab to require and encourage thinking sends conflicting messages. This is similar to a waiter

---

[1] This example was taken from one of the interviews described in Chapter 4. For more information, see the section on "Ostensible Purpose" in that chapter.

wearing sweatpants and a T-shirt telling you the restaurant is high-class and formal. Given conflicting messages, students may fall back on what they are used to, and go into their traditional laboratory frame.

Adding questions such as "Do these trends make sense?" to the lab manual and requiring the students to explain why things happen may cause some students to go into sense-making more frequently.[1] This occurred for Veronica and Carl (see Chapter 5); however, they were very successful students. This change was not productive for most other students in the lab. In addition, lab occurred directly after tutorial, where students were guided through an inquiry process requiring sense-making. Students had other experiences in physics class where they were used to making sense which made it easier for them to enter a sense-making frame.

For the same reason that adding thinking questions to a traditional lab may fail, adding one scientific community lab to a string of traditional labs is likely to fail. Having found a behavior that works, students will stick with it. Switching lab types throughout the semester will send conflicting messages to students, and they are likely to rely on traditional laboratory behavior.

Measurement Concepts

As discussed in Chapter 3, there are two measurement concepts that we have been unable to find mention of elsewhere, but which are critical for students' understanding of measurement and uncertainty and for teachers' interpretation of students' understanding. If not included in the laboratory curriculum, we recommend that they at least be included in the training for laboratory instructors. These concepts are *internal and external variation* and *predictive and descriptive questions*.

Internal and External Variation

Internal and external variation differ from each other in their effect on the results of the measurement and their cause. External variation (often called "human error" by students) comes from the measurement method and has a detrimental effect on the experiment's results. Internal variation comes from something internal to the system being measured, and has a neutral effect. (It may be what the experiment is measuring.) For example, suppose a sleeping bag manufacturer wanted to claim that their sleeping bag was warmer, and designed a test to prove this. They test five of each type of sleeping bags by heating water to boiling, pouring it into a bottle, putting the bottle inside the sleeping bag inside a refrigerated meat locker, and measuring the temperature of the water after 6 hours. External variation could come from differences in the temperature of the bottle before the water is poured in – differences for which the researcher is responsible. Internal variation could come from differing amounts of stuffing in the sleeping bag – differences for which the system manufacturer is responsible.

This difference is critical especially for the types of questions mentioned in the previous section, for example, "If you use the device and method described below, how well can you measure a food's weight?" Such questions ask students to reduce external variation while measuring internal variation. Students who conflate the two ideas will be unsure of what to measure, and may treat all variation as "error".

Predictive and Descriptive Questions

A predictive question asks what may happen in the future, and a descriptive question asks what already happened. Such distinction seems trivial, but typical lab questions such as "Is momentum conserved when two pucks collide?" could be interpreted either way. Students who fail to look at the spread in values and report only what actually occurred may be answering a descriptive question, while the laboratory instructor is looking for a predictive answer. For the sleeping bag example, supposed the manufacturer claimed their sleeping bag was warmer. If it won the test 3 out of 5 times, one perhaps

---

[1] See Appendix A for the cookbook+explanation lab handout, an example of a traditional cookbook lab with added "explain" questions.

could claim that those three sleeping bags tested were warmer than their rivals. But one would not be able to recommend any sleeping bag from that brand to a prospective customer. The question "how well can you measure a food's weight?" is asking for a predictive, not descriptive answer, but this may not be clear to students.

These two ideas can lead to miscommunication between instructors and students in the laboratory, and lead teachers to attempt interventions which will be unlikely to succeed. A student answering a descriptive question will unable to make sense of a request to look at the range in the data. Instructors and students should be aware of these differences, and questions asked in the laboratory should be unambiguous.

Using research tools on other lab designs

We use several different qualitative and quantitative tools in this dissertation to study students' understanding of measurement and uncertainty and to evaluate the scientific community labs. Some of these tools may be directly applied to other lab designs and student populations, and others may need to be adapted to work in new situations. In this section we review the specific research tools, their purpose, and how they can be adapted.

Anonymous Survey

One very simple way to evaluate students' perceptions of a curriculum is through the four questions on the anonymous survey introduced in Chapter 4. These questions ask what the student learned in lab that will be useful to them in the future, what the pedagogical purpose of lab is, what they would change about lab, and how useful lab was.[1] This survey was useful for comparing two different laboratory curricula. It also may be useful for evaluating different instructors using the same curriculum, observing whether different instructors are able to more clearly communicate the goals and future uses of the lab to the students.

Cognitive Behavior Mode

By analyzing students' small group conversations during lab we can determine whether the group is engaging in sense-making, logistics, or off-task behavior.[2] This coding scheme can be used to evaluate a laboratory design by studying what kinds of behavior the design encourages. It is time intensive to complete, requiring videotaping, transcribing, and then coding. It may be possible to code students' behavior in real time while observing them during lab or while watching the videotape without needing to transcribe. This would require careful training and reliability testing, but it would make this type of coding more feasible.

While working in the laboratory, students are frequently monitoring and evaluating their behavior using metacognitive statements such as "I don't get this" or "I don't understand why that didn't work". Often such metacognitive statements do not have any effect on students' behavior: they continue in the same mode which led to their confusion. In one traditional cookbook laboratory, the coders identified 15 metacognitive statements in one hour, but none of them caused the students to change cognitive modes. Students may explicitly state that they do not understand something but then do nothing about it because they expect and are resigned to confusion during the laboratory. Coding for metacognition allows one to evaluate whether the design of the laboratory encourages students to monitor their behavior and change to a more productive mode. It is very difficult to reliably identify metacognitive statements. Even using a transcript and the videotape, the two coders did not always agree on which statements were metacognitive, so in our study only statements originally identified by both coders were taken as

---

[1] The full anonymous survey is included in Appendix C.

[2] The Cognitive Behavior Mode coding scheme is described in Chapter 5, and an example of a coded transcript is included in Appendix E.

metacognitive. Currently metacognitive coding requires transcribed video and two coders, making it extremely time consuming, though interesting and valuable research.

Data Set Comparison

We used the method of knowledge analysis[1] to find the ideas underlying students' decisions in the laboratory and then to develop a scheme for analyzing students' arguments concerning data set comparison. The analysis scheme uses a flow-chart representation to organize students' arguments according to complexity. This representation allows one to study how students' arguments change as a result of instruction.[2] It could also be used to investigate whether a certain environment or lab context elicits more or less complex arguments from students.

This is another time intensive method, requiring one to identify useful video clips, transcribe and analyze the clips. Obviously, it requires useful video clips: students must be talking about comparing two data sets, making clear arguments and defending their conclusions. This may not happen in certain lab designs – we have no record of it ever occurring in traditional introductory physics labs at the University of Maryland. The scheme could also be used to analyze written student arguments, if they are detailed enough.

Measurement Survey

In Chapter 7 we describe the development of a multiple choice survey for studying students' ideas about gathering, analyzing, and comparing data. This survey can be used to study the effect of a certain instructional technique on large numbers of students, requiring much less time than the previous research tools. It is designed to be administered electronically, which controls the order of the questions and the answer choices displayed.[3] (This also makes it very easy to administer and analyze.) Students' responses to each individual question or their amount of reasoning from the set or point paradigm can be evaluated. When doing the latter, a histogram showing the number of students with certain amounts of set reasoning and a scatter plot showing the amount of set reasoning versus point reasoning may be useful.

Because of time limitations, this survey contains a small number of questions (8) carefully chosen for the population of the algebra-based introductory physics course at the University of Maryland.[4] The survey questions may be too difficult or too easy for other student populations, and may not distinguish well between students. In this case, one may need to use more suitable questions from other surveys (Buffler et al, 2001; Leach et al., 1998; Lubben and Millar, 1996). The answers for the multiple-choice survey were selected from our population's free responses, so when switching populations it may be necessary to first administer a preliminary free-response version to see if the new population uses the same answers.

References

Buffler, A., Allie, S., Lubben, F., and Campbell, B. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *Int. J. Sci. Educ., 23*(11), 1137-1156.

---

[1] See Chapter 6 for more information on Knowledge Analysis.

[2] Selected transcripts and flow charts are included in Appendix E.

[3] The multiple choice survey questions, results, and previous versions of the same survey are included in Appendix F.

[4] See Chapter 3 for more information on the student population of the algebra-based course.

Leach, J., Millar, R., Ryder, J., Sere, M-G., Hammelev, D., Niedderer, H., and Tselfes, V. (1998). Students' images of science as they relate to labwork learning. *European project LABWORK IN SCIENCE EDUCATION* (Working Paper 4 (Targeted Socio-Economic Research Programme Project PL 95-2005).

Lubben, F., Millar, R. (1996). Children's ideas about the reliability of experimental data. *Int. J. Sci. Educ., 18*(8), 955-968.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In Schoenfeld, A.H., *Cognitive Science and Mathematics Education* (189-215). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chapter 9: Summary and Future Research

Introduction

The main goals of this dissertation concern both research and instruction: to study how students understand measurement and uncertainty and to design a curriculum which teaches this. We argue that students must be in a frame where they will use productive resources to build an understanding of the underlying concepts and then apply those concepts to data analysis. In this chapter we first summarize the details of this process and then propose further research which could be done to build on this project.

Dissertation Summary

For this final summary we present a documentary[1] of a stereotyped, idealized student we shall call Demeter and her experiences in the meta-learning course and scientific community labs. Demeter's story portrays a student's view of this project, which we follow with a researcher's view.

We first meet Demeter as she enters the introductory physics classroom with all her expectations about how the course will proceed. We then follow her progress throughout the course using the research tools presented in previous chapters. The chapters where each idea is discussed are noted as numbers in parentheses.

Demeter's story

Demeter is a Biology Major in her junior year, hoping to do research of some sort upon her graduation (Chapter 3). She has already taken physics in high school, two semester of calculus, and more introductory science laboratories than she can count. Most of them were rather useless and unconnected to the course, though sometimes it was fun to play with the equipment. Lab was a chance to see the lecture stuff in action, if everything went well and the lecture and lab actually coincided (Chapter 4). She had learned that one should repeat each experiment at least twice, usually five times, and average the results – this makes the results more accurate, reduces human error. Most experiments result in an average (Chapter 7). She also remembers something about standard deviation and chi-squared, though those ideas never really made sense. It's the lecture stuff that one tries to make sense of, if possible. The lab stuff one just learns to do (Chapter 4).

The course begins with a whole bunch of surveys they have to take. Most of the questions are strange, at least for a physics course. But at least they get extra credit for taking them. They also get points for answering questions in lecture with a keypad (Chapter 3). It's kind of odd, the professor seems to want people to respond and talk during lecture. Maybe the keypad is just a way of taking attendance.

The first lab goes by. Demeter meets the other students she'll be working with. Lab is spent mostly watching the TA, who is waiting for students to say something, anything. The TA mentions that their grade is partly based on participation, and then more students speak up. This lab seems a little different: they won't be given a lab manual, they'll write up their lab report as a group, and they'll have to make a presentation to the whole class at the end of the lab (Chapter 3).

The second lab leaves Demeter exhausted and angry. How in the world can anyone expect them to finish all this in two hours? They have barely 20 minutes to complete the experiment, and even less time to plan it. She felt like an idiot, standing there in front of everyone presenting a half-completed

---

[1] A note on frames: this chapter could serve as an example of a frame switch. Notice the difference between the documentary style of the summary section and the scientific paper style of the other sections. One difference is the use of conjunctions in the documentary style, which makes it more conversational. Also the documentary style is organized sequentially in time, instead of grouping logical ideas.

experiment (Chapter 4). At least their results came out nicely – the two averages for changing the length of the pendulum were clearly different, and the two averages for changing the mass were almost identical (Chapter 6). Some groups weren't so lucky, and then the TA bugged them about it, asking "is that significant?" How is one supposed to know significance, anyway, without running some kind of analysis software package, and the lab is barely equipped with enough stopwatches, much less any technical equipment.

A few labs go by. Demeter notices that, because of the time crunch, she and her group members have to evaluate their actions, deciding if their current activity is really worth spending time on, and asking if they're making progress (Chapter 5). She is actually using some of the stuff from lecture to make sense of what's going on in lab – they have to understand what they're doing to plan their experiment and present it to the class. If they don't explain it clearly, the other students don't understand, and then they get asked a lot of questions after their presentation.

Recently lab seems to be very focused on range and range overlap, but she's still not sure about how to tell if two averages are significantly different or not. The class has come up with many different arguments claiming two things are the same or different (Chapter 6), but how do you know which is the best to use? The TA never gives any answers, just tells them to use the one which is most convincing. Some students start debating two ideas; Demeter thinks it's best to just take so much data that it's clear whether they're the same or different. Unfortunately, they don't have time for that in most of the labs.

Demeter gets her lab quiz back. She didn't do so well, so she reads over the solutions and talks to the TA. During the next lab, she realizes that it helps to think about all of this stuff as if she were gambling, as if she would win money by predicting what the next measurement would be. The average is like the best bet, but the next measurement could lie anywhere around that within a certain range. The size of the range says how good the average is. Another thing she realizes– the evaluation isn't supposed to be a summary. It's supposed to be like you're telling your parents why they are wrong and they should let you go to a party, or telling your brother why he should let you use his car. Why the TA didn't tell them that in the beginning, Demeter doesn't know.

During the class discussion Demeter realizes that another group of students did things very similar to her group, but her group got a different average and a smaller range. They both measured the speed of the marble after rolling down a ramp, but her group put two small bumps in the track so they could hear the marble hit them and use the sound to know when to start and stop the stopwatch. Demeter argues that her group got a smaller range so their method is better, but the other group argues that the bump may have slowed down the marble and made all their times longer, so their average was too large.[1] Both groups ask the TA, who doesn't answer the question (of course) but instead asks the rest of the class. After getting nowhere, the TA ends with some excuse about how some of these things do not have a clear answer.

It's the end of the semester, and another round of surveys. Demeter vaguely remembers the online lab survey from the beginning of the semester (Chapter 7), because she had such a hard time logging in to WebCT. This time it's rather frustrating because all the questions are the same: they all basically ask "does the range overlap?" There should be a checkbox for "see previous answer". Oh well, she'll take any extra credit she can get.

Narrator's comments

Having described the project from a student's point of view, we now switch to the researcher's viewpoint and summarize the goals, successes, and challenges of this project. Informal observations of the cookbook style physics lab have shown that students seem to do very little thinking during lab and tend to blindly following instructions. Most students seem to leave the course with little or no understanding of measurement or of uncertainty, knowing little more than that standard deviation is

---

[1] These two ideas are similar to systematic and random uncertainty. For students to be able to debate whether systematic uncertainty or random uncertainty is better is a goal of the scientific community lab.

connected to uncertainty. The ideas of uncertainty and measurement will be useful for students in the future, and especially for the biology majors served by the algebra-based physics course. Thus, we set out to design a physics laboratory curriculum which would leave students with an understanding of the basics of measurement and uncertainty.

We designed the scientific community lab to create an environment where students would develop measurement and data analysis strategies which were convincing to others. To encourage students to build an understanding of these ideas and not just memorize tools, each section was allowed to use only the tools which they could define and explain satisfactorily to their fellow students. This meant that standard deviation and other more mathematical tools of uncertainty analysis were rarely, if ever, used (only one section out of 8). Instead, students built upon the basic ideas of range, average, and percent difference.

To determine whether the lab design successfully created a productive environment, we interviewed and surveyed students for their ideas about the purpose and usefulness of the lab. Different parts of the frame setting were more or less successful. The scientific community labs were only partly successful at communicating the main goal of the labs: for students to learn how to gather and analyze data. Students in the traditional laboratory saw the primary goal of the laboratory as learning physics concepts (95%), and some students remarked that physics concepts will be useful in their future (22%). In contrast, students in the scientific community lab identified learning problem solving, experimental, and other skills as a purpose of the laboratory (64%), and many claimed that these would be useful for their future (86%). Only 8% of students mentioned learning to analyze data. We managed to convince students of the usefulness of learning skills in addition to concepts. But it was difficult to effect such a large change, from learning physics concepts to building skills for analyzing data.

It also was difficult to get students to critique each other's lab designs and to appreciate critiques of their own lab designs. Frequently, at the end of lab, students would ask for the "right" answer. They failed to see, or did not believe, that the purpose of lab was not to get the right answer, (if such an answer even existed) it was to make the strongest argument. Further research is needed to determine how to fully develop a scientific community. We discuss this in the next section.

Another way to test if the lab design created a productive environment for the students is to observe student's behavior during lab. Students in the traditional laboratory spend little time sense-making: 4% of the total time in the two cookbook labs studied. Students in the scientific community labs spend significantly more time sense-making, about 20%. They are also more likely to monitor their actions in a way that causes them to change to a more productive mode.

Once students are sense-making, we want them to build an understanding of the concepts underlying measurement and uncertainty, and then use these concepts to analyze data. During lab students developed complex arguments about comparing two sets of data, so we used a flow-chart diagram to illustrate the student's underlying argument structure. We then developed a survey to test student's ideas about gathering, combining, and comparing data. After completing the scientific community labs, students were more likely to use information about the spread in the data to compare the two data sets, instead of just comparing the averages (12% before to 43% after).

It appears that students leave lab with an understanding of the basic concepts underlying uncertainty analysis: what information comes from taking multiple measurements, how to estimate the uncertainty in a measurement, and how to compare two data sets. With this understanding they are now able to learn the mathematical tools of uncertainty analysis and make sense of why they are useful and how to productively apply them. Students who go into the health services or biology research will need to interpret numbers such as standard deviation, chi-squared, and R-squared. The development and evaluation of a lab designed to teach this in the second semester physics course is one area for future research. Other areas are discussed in the next section.

Future Research

As with many scientific studies, this research may have raised more questions than it answered. Some of them are obvious: research is needed to disseminate the existing curriculum and research tools (some of which is mentioned in the previous chapter), and research is needed to develop a second semester lab which builds on the scientific community labs. There are also larger questions about the broader issues. We discuss these questions, some of which head off into related fields, in the remainder of this chapter.

Building a scientific community

The scientific community labs were designed to create, as much as possible, a scientific community in the student laboratory. Students were given a research question, but then they designed their own method, formed their own conclusions, and defended their conclusions to the group. Each section defined their own tools and terms for communicating among groups. Students were supposed to critique other groups' methods and work together to determine the best method and conclusion. This last goal was very difficult to realize, perhaps for several different reasons.

It may be that students do not have well developed critiquing skills. Especially in the beginning of lab, students struggled to write the evaluation section of their lab reports and to answer the evaluation questions on the lab quiz. They would often simply summarize the method or claim that the students should have taken more data. This problem diminished as time went by and students began to understand what the TA wanted in an evaluation, but several students in each section never managed to successfully evaluate a lab. Were these students unable to critique in any scientific setting? Were they unable to critique even in a social setting? It may be necessary to study how to teach students to critique.

One student, while talking to the professor, clearly stated another possible reason for the problem. The professor asked how lab was going, and she replied that things were pretty good, except for the class discussion which was kind of boring. The professor questioned whether she was paying attention, asking questions, and critiquing as she was supposed to be doing. The student replied (her response is paraphrased here) "Oh no – I learned long ago, if you want to have friends, you never tell someone else they've done something wrong."

Critiquing someone seems to be socially frowned upon in undergraduate student culture, and perhaps even in graduate student culture. Two professors who frequently work together were helping to run TA training for graduate students. One activity involved first watching a clip of students working and then brainstorming what the students were thinking and possible interventions. The two professors started arguing over an interpretation. As the graduate students were leaving, some of them commented that those professors must really hate each other. The students reasoned that if the professors were friends, they would not disagree and argue with each other.

If social acceptance forbids critiquing, then what happens for graduate students to make the transition to professors? Do they stop caring about social acceptance? Do their social skills degrade? We do not want students to be rude or insulting, but we do want them to feel free to disagree with each other and point out possible mistakes. For the lab to create an environment where students are encouraged to critique others, this issue must be addressed.

Gender/Cultural Research

The attempt create an environment where students are in a productive frame is basically an attempt to create a culture. We want to take students, with all their different cultural backgrounds, and meld them together into a scientific community. How do students from different cultures react to this? Is it easier for students from one culture, and harder for students from another? There may be aspects of certain cultures we could borrow to create the lab culture, making it easier to communicate the goals of lab.

These same questions could be asked about gender differences. Perhaps female students are discouraged from critiquing others, and male students are discouraged from working together in the majority culture. Certain aspects of the lab may be more rare or more common for male or female students, more acceptable behavior or more unacceptable behavior. Previous research has show that cooperative learning may help to reduce the gap between male and female students' achievement in science (e.g. Haukoos and Penick, 1983; Lazarowitz et al., 1985). Does this occur in the scientific community labs? These and other questions could be studied about different students' experiences in the lab.

Convincing Motivation

As part of the scientific community, students are encouraged to "be convincing" and are graded for how convincing their lab reports are. This is to support a productive frame, for the skills and tools of uncertainty analysis are needed to be convincing. We have some evidence that this is productive, from students asking "is this convincing" while working in their group during lab, and from students writing "this argument is not convincing" on their evaluation section of the lab quiz. Further research could be done on how productive this frame-setting device is, and on how to make it even more productive.

In summary, there exist many more research questions about how to create a mini-culture in the physics laboratory which provides the right environment for students to build an understanding of measurement and uncertainty, as well as other questions about what exactly students are building an understanding of and what it means to build an understanding. We have attempted to answer parts of these questions in this dissertation, but much still remains unanswered.

References

Haukoos, G. D., and Penick, J.E. (1983). The influence of classroom climate on science process and content achievement of community college students. *Journal of Research in Science Teaching, 20*(7), 629-639.

Lazarowitz, R., Baird, J.H., Hertz-Lazarowitz, R., and Jenkins, J. (1985). The effects of modified jigsaw on achievement, classroom social climate, and self-esteem in high school science classes. In Slavin, R., Sharan, S., Kagan, S., Hertz-Lazarowitz, R., Webb, C., and Schmuck, R., *Learning to cooperate, cooperating to learn* (231-253). New York: Plenum.