The Pennsylvania State University

The Graduate School

College of Education

SECONDARY SCHOOL PHYSICS TEACHERS' CONCEPTIONS OF SCIENTIFIC EVIDENCE: A COLLECTIVE CASE STUDY

A Thesis in
Curriculum and Instruction
by

Joseph A. Taylor

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2001

We approve the thesis of Joseph A. Taylor.

Date of Signature

Thomas M. Dana
Associate Professor of Education
Thesis Advisor
Chair of Committee

Vincent N. Lunetta
Professor of Education

Carla Zembal-Saul
Assistant Professor of Education

Renata S. Engel
Associate Professor of Engineering Graphics and Engineering Science & Mechanics

Patrick W. Shannon
Professor of Education
Coordinator for Graduate Program in Curriculum and Instruction

ABSTRACT

Joseph A. Taylor

Thesis Advisor: Thomas M. Dana

Engaging secondary school students in inquiry-oriented tasks similar to the scholarly activities of scientists has been recommended as a way to improve scientific literacy. Two frequently recommended tasks are students' design of original experiments and students' evaluation of scientific evidence and conclusions. Science education scholars have suggested that teachers must possess well-developed conceptions of scientific evidence to guide students as they engage in these tasks. Yet, little is known about teachers' conceptions of scientific evidence.

The principal aim of this study, therefore, is to describe the nature of one prospective and two practicing physics teachers' conceptions of scientific evidence. More specifically, the following research questions guided this study: (1) What types of issues related to the measurement reliability and experimental validity of scientific evidence do the participant-teachers think about when designing experiments? (2) When presented with hypothetical scenarios that describe unsound experimental procedures or poorly supported conclusions (or both), what concerns will the teacher-participants raise? And (3) When the teacher-participants' responses to parallel research prompts are compared across protocols, what similarities and differences exist?

The nature of the teacher-participants’ conceptions was derived from an analysis of data collected from research prompts such as interviews and hand written artifacts. In these research prompts, the teachers “thought aloud” while designing experiments and evaluating student–collected evidence presented in hypothetical classroom scenarios. The data from this study suggest that the three teachers, while contemplating the reliability and validity of scientific evidence, frequently used their conceptions of evidence in conjunction with specific subject matter conceptions. The data also indicates that the relationship between subject matter knowledge and conceptions of evidence was more pronounced for some conceptions of evidence (e.g., measurement validity, controlled experimentation, statistical significance) than for others.

Suggestions for future research included conducting similar studies in other physics content areas as well as other scientific disciplines. Implications for science teacher education suggest that science and science methods courses encourage teachers to conduct original research and to construct and present evidence-based arguments from this research for peer review and critique.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to extend sincere thanks to my thesis advisor Thomas M. Dana for his time, guidance, and expertise. This thesis is compelling “evidence” of our effective collegial work and close friendship. I would also like to thank Vincent Lunetta, Carla Zembal-Saul, Renata Engel, and Susan Land for providing insightful direction and feedback while serving on my doctoral committee.

The encouragement and support given to me by my parents, extended family, and close friends was greatly appreciated and kept me going through the tough spots. A special thanks goes out to my wife Sara for her moral and financial support throughout my doctoral program and for her patience as she listened to daily updates on the progress of this thesis.

Lastly, I would like to dedicate this thesis in memory of my mother-in-law Jennie Freeman. Jennie was a special person to me and was my wife’s best friend. I know that she is happy for both of us and shares in this accomplishment.

CHAPTER 1

SETTING THE STAGE

Concerns about the state of science education in schools have been highly visible in the research and policy literatures for decades (e.g., DeBoer, 1991). A historically popular notion related to improving science education is that science teachers should strive to narrow the gap between scientists' science and the image of science presented in K-16 science classrooms (DeBoer, 1991). The rationale seems to have been that being scientifically literate means more than just knowing factual science information. In fact, recent policy documents such as *Benchmarks for Scientific Literacy* (AAAS, 1993) and the *National Science Education Standards* (NRC, 1996) help refine a national goal for science instruction that emphasizes active learning of science content through scientific inquiry.

Yet, the task of engaging students in scientific inquiry is fraught with additional concerns. What does it mean to engage in scientific inquiry? What do students understand about scientific experimentation, data collection, and analysis? While student engagement in scientific inquiry could be a useful tool to accomplish goals of scientific literacy, successful inquiry requires a student to learn a whole new set of conceptions related to experimentation and the uses of scientific evidence.

Many science education scholars report research findings about students' alternative conceptions of the nature of scientific knowledge construction (e.g., Aikenhead, 1987; Lederman & O'Malley, 1990; Rubba & Anderson, 1978). Specifically, the research literature reported that many K-16 science students inadequately distinguish between theory and evidence (Allen, Statkiewitz, & Donovan, 1983; Lawson, Clark, Cramer-Meldrum, Falconer, Sequist, & Kwon 2000), view scientific knowledge as static and purely objective (Aguirre, Haggerty, & Linder, 1990; Desautels & Larochelle, 1998; Kitchener, 1983; Kitchener & King, 1981), and fail to recognize the possibility of fallible interpretation in the construction of scientific knowledge (Abell & Smith, 1994; Aikenhead & Ryan, 1989; Carey, Evans, Honda, Jay, & Unger, 1989). Some science educators attributed these naïve notions, at least in part, to common and insufficient instructional practices (especially in physics education) such as engaging students in verification-type laboratory investigations and ritualistic problem solving (e.g., Driver, Newton, & Osborne, 2000; McDermott, 1990a; Woolnough & Allsop, 1985).

Hoping to engage students in activities that more closely simulate the scholarly activities of scientists, many scholars and researchers in science education have recommended the infusion of science-specific inquiry-oriented tasks into K-16 science instruction (e.g., DeBoer, 1991; Dewey, 1933; Herron, 1971; NRC, 1996; Schwab, 1966; Suchman, 1966; Tamir, 1983). Although slightly different in focus, the influence of these recommendations is evident in historic science curriculum design efforts (e.g., Biological Sciences Curriculum Study - BSCS, Physical Sciences Study Committee - PSSC) and in more contemporary teacher resource materials such as *Invitations to Science Inquiry* (Liem, 1987) and *Physics by Inquiry* (McDermott et al., 1996).

Two student tasks that are often emphasized in these and similar curriculum materials are students’ design of experiments and evaluation of evidence and conclusions. The emphasis on these tasks is consistent with the vision of inquiry-based science learning described by scholars in science education (e.g., Driver, Leach, Millar, & Scott 1996; Lawson, 1995; Lemke, 1990). It is also apparent in documents like the *National Science Education Standards* (NRC, 1996), which has envisioned science learning as including student involvement in “planning investigations” (p. 23) as well as “weighing the evidence, and examining the logic so as to decide which explanations and models are best” (p. 175).

In the last decade, however, there has been a shift in some science educators’ opinions about students’ thinking while engaged in these types of tasks. Some have theorized that students’ ideas about experimentation and scientific evidence are more than just “process skills”; they have claimed that these ideas are concepts in their own right (e.g., Gott & Duggan, 1996; Lubben & Millar, 1996; Millar & Driver, 1987; NRC, 1996). Specifically, Gott and Duggan (1995) suggested that students call upon a distinct set of conceptions regarding the reliability and validity of scientific evidence when designing and conducting experiments and when evaluating the resulting evidence and conclusions. In a broad sense, Gott and Duggan related reliability and validity to the following questions respectively: can the data be believed and can the data answer the question(s)? Extending Gott and Duggan's line of reasoning would suggest that students must possess certain understandings regarding the reliability and validity of scientific evidence before they can meaningfully engage in the inquiry-oriented tasks recommended in the literature.

How and where, then, do students develop these appropriate understandings of scientific evidence? A number of science educators have emphasized the crucial role of the science teacher as one who can provide the guidance necessary to help students develop more appropriate conceptions of the reliability and validity of scientific evidence. For instance, *Science for All Americans* (AAAS, 1989) suggested that science educators must guide students in "collecting, sorting, and analyzing evidence, and in building arguments based on it" (p.201). Although many science educators have suggested that teachers "support" students as they attend to issues related to scientific evidence, few have been very specific about what this support entails or what conceptual understandings teachers need to have in order to provide it. Newton, Driver, and Osborne (1999) suggested that (procedurally) science teachers might provide support by asking open-ended questions that stimulate students to think critically about their data or conclusions. Huber and Moore (2001) have indicated that such questions might be as straightforward as “how will you explain or defend that conclusion in your class presentation?”

Although broad questions such as these might be effective in initiating some degree of student discourse, it is unlikely that such questions will encourage students to thoroughly evaluate the reliability or validity of their data or conclusions without further probing by the teacher. As students respond to broad questions, science teachers must be able to follow these responses with a more focused and detailed line of questioning, specific to the inquiry at hand. What is not clear, however, is the nature of the understandings that science teachers must possess in order to formulate these questions that are likely to stimulate critical thinking about scientific evidence.

Purpose of this Study

This study is designed to investigate the aforementioned understandings about scientific evidence that teachers must possess. To accomplish this, three physics teachers were studied. I chose to study physics teachers as I have a background in this area of science and I wanted to focus on conceptions of scientific evidence in a singular science domain. The principal aim of studying the three teachers was to describe their conceptions of the measurement reliability and experimental validity of scientific evidence.

Research Questions

An overarching question guided the design and execution of this study: What is the nature of prospective and practicing physics teachers' conceptions of the measurement reliability and experimental validity of scientific evidence? More specifically, the following research sub-questions provided a greater focus:

a) What types of issues related to the measurement reliability and experimental validity of scientific evidence do the participant-teachers think about when designing experiments?

b) When presented with hypothetical scenarios that describe unsound experimental procedures or poorly supported conclusions (or both), what concerns will participant-teachers raise?

c) When the participants-teachers' responses to parallel research prompts are compared across protocols, what similarities and differences exist?

Research on Conceptions of the Measurement Reliability and Experimental Validity of Scientific Evidence

Historically, the definitions of “reliability” and “validity” have depended on their context. This literature review will focus on the more specific issues of *measurement reliability* and *experimental validity*. Wiersma (1995), though mainly concerned with educational research methods, described measurement reliability and experimental validity in general terms that can be applied to school-based science investigations as well. He stated that measurement reliability pertains to the consistency with which the chosen instrument measures whatever it measures (p. 309). On the other hand, experimental validity pertains to questions such as:

- To what contexts can the results be generalized (p. 112)?
- Does the experimental design support the collection of interpretable data by imposing adequate control of extraneous variables (p. 112)?
- Do the instruments measure what they are designed to measure (p. 311)?

This literature review is divided into three subsections. The first deals solely with research into K-16 students’ conceptions of the measurement reliability of scientific evidence. The second describes studies of K-16 students’ conceptions of the experimental validity of scientific evidence. Finally, the last section examines research into prospective and practicing science teachers' conceptions of both the measurement reliability and experimental validity of scientific evidence.

K-16 Students' Conceptions of Measurement Reliability

Research with elementary school children suggests that the need for repeated measurements is not always properly understood (Carey, Evans, Honda, Jay, & Unger, 1989; Schauble, 1996; Schauble, Klopfer, & Raghavan, 1991; Varelas, 1997). Schauble (1996), for instance, found that many of the fifth and sixth grade students studied did not plan to replicate measurements; they believed the initial measurement had to be the “correct” value. When students inadvertently repeated measurements, they were disturbed with even slightly dissimilar numeric values. Similarly, Varelas (1997) found that many third grade and fourth grade students articulated the idealized viewpoint that equivalent experimental conditions should always yield identical results. Students who recognized that random experimental error might occur (under equivalent experimental conditions) were not as concerned about the dissimilar measurements as the students who held the idealized view, but Varelas observed that few students integrated ideas from both viewpoints. In similar studies, others observed this apparent dichotomy of ideas as well (e.g., Duschl, 1990; Holton, 1988).

Although many middle and high school science students claimed that repeating measurements is necessary to establish reliability, few articulated a meaningful rationale for doing so (Foulds, Gott, & Feasey, 1992; Gott & Duggan, 1995; Lubben & Millar, 1996). Likewise, a superficial understanding of the rationale behind repeated trials was observed in post-secondary students (e.g., Allie, Buffler, Kaunda, Campbell, & Lubben, 1998; Sere, Journeaux, & Larcher, 1993). For example, Allie, Buffler, Kaunda, Campbell, and Lubben (1998) found that many undergraduate physics students saw the purpose of repeating measurements solely as a way of producing more than one data point so that a mean could be calculated.

Lubben and Millar (1996) conducted a large-scale study of the conceptions of data reliability held by year 7, year 9, and year 11 science students. The findings of their study influenced much of the subsequent research in this field and will be referred to often in this review. They asked the students to suggest why repeated measurements might be essential to the collection of reliable evidence. The most common conceptions, as well as the percentage of each age group holding those conceptions, are shown in Table 1.1.

Table 1.1. Lubben and Millar’s Findings: Conceptions of Repeated Trials

Common conceptions of the rationale behind repeated trials	Year 7 students (n=223)	Year 9 students (n=188)	Year 11 students (n=122)
A. Repeated trials are not needed	6%	3%	2%
B. Repeats required for confirmation of “correct” answer	35%	19%	13%
C. Multiple repeats are needed so that an average can be calculated	22%	13%	7%
D. As a part of scientific routine	11%	13%	6%
E. To allow for scatter (variance)	18%	50%	71%

Source: Adapted from Lubben and Millar (1996)

Building on their repeated trials research and other related studies with 11-16 year-olds, Lubben and Millar (1996) developed a model that establishes a hierarchy of common conceptions regarding the reliability of experimental data (see Table 1.2).

Table 1.2. Levels of students' understanding of the collection and evaluation of experimental data

Level	View of the process of measuring	How to evaluate your result	What to do with readings that differ greatly from most of the others
A	Measurement is straightforward; measure once and you will get the right value	Not an issue. A measurement is correct.	Not an issue.
B	Measure once and take this as the right value. Repeating will lead to a different result, but any result is likely to be as good as any other, so there is no point in repeating.	Unless something has obviously gone wrong, a measurement is correct. In familiar contexts, it should be close to what you expect for the quantity measured.	Not an issue.
C	If you have adequate equipment and use it carefully, your measurement will be right. Make a few trial measurements to become familiar with the equipment; then take the measurement you want.	Unless something has obviously gone wrong, a measurement taken after a few trial runs will be correct. In familiar contexts, it should be close to what you would expect for the quantity measured.	Ignore. (If a 'trial' reading is different, this is due to lack of practice; if 'final reading is different, this is the result of practice.)
D	If you have adequate equipment and use it carefully, your measurement will be right. Repeat measurements in order to get the same result twice.	Getting the same value twice shows you have measured carefully enough.	Ignore.
E	You should repeat a measurement and take the average. But repeating exactly the same measurement will lead to the same value; so change the conditions a little each time.	Variation is to be expected. Not an issue.	Variation is to be expected. Include all values in calculating an average.
F	Careful measurements may get close to the right value of the quantity you are measuring but you can never be sure you have found it. Taking an average of several measurements allows for this.	Cannot be evaluated from 'inside.' Only method is to check with an authority source (the teacher, or a textbook or databook).	Including them when calculating the average will take care of them. In fact, this is why we use the average.
G	As above.	The spread of the measurements is an indication.	As above
H	As above.	As above	Reject anomalous results before taking an average.

Source: Adapted from Lubben and Millar (1996)

In their model, Lubben and Millar (1996) recognized that important questions follow naturally from the initial question: why repeat measurements? One such question concerns how to deal with the situation when one or more measurement values is significantly different than most of the others. Other scholars and researchers have also addressed the question of how students deal with anomalies or “outliers” (e.g., Chinn & Brewer, 1993; Joshua & Dupin, 1987; Kuhn, 1989).

Chinn and Brewer (1993) reviewed much of the existing literature regarding students’ responses to anomalous data and suggested that students’ strategies (see Table 1.3) for

Table 1.3. Common student strategies for dealing with anomalous data

Student Response to Anomalous Evidence	Notes
Reject the anomaly without justification	The anomaly is simply ignored (in an effort to preserve existing theories).
Reject the anomaly with justification	The rationale for rejecting the anomaly may include issues such as the evidence is an artifact of random variation, the evidence is a hoax (the experiment has been ‘rigged’) or the experimental procedure used was faulty.
Exclude the anomaly from the domain of the current theory/hypothesis.	Anomaly is excluded from consideration by virtue of being outside the domain of the current theory.
Hold the anomaly in abeyance	Procrastinate in dealing with the anomaly
Reinterpret the anomaly in such a way as to preserve existing theories.	_
Reinterpret the anomaly making only peripheral changes to existing theories	_
Accept the anomaly and change the existing theory	Rarely occurs if few anomalies conflict with existing theory/hypothesis.

Source: Adapted from Chinn and Brewer (1993) dealing with anomalies tend to fall into seven categories. Most of these strategies allow students to preserve existing theories or hypotheses. This finding is consistent with those of other researchers who documented the tenacity of students' pre-existing theories (e.g., Champagne, Klopfer, and Gunstone, 1982; Clement, 1982; Lewis and Linn, 1994).

Nevertheless, there are instances when the rejection of an anomaly is appropriate. Knowing when this rejection is warranted and knowing how to deal with the remaining data are important parts of a well-developed concept of the reliability of scientific evidence, and it is here that the work of Chinn and Brewer (1993) intersects that of Lubben and Millar (1996). Chinn and Brewer observed that students' rationales for anomaly rejection often included issues such as accounting for random variation and experimental design. Lubben and Millar asked the students to respond to a set of data that included an anomaly (see Tables 1.4 and 1.5), and few students, regardless of age group, even recognized the anomaly. In addition, a considerable number of students either included the anomaly in their data analysis or chose to represent the data set with an inappropriate measure of central tendency.

Inappropriate measures of central tendency were also preferred by many of the students who did not recognize the anomaly.

Table 1.4. Percentage of students recognizing the anomaly

Age Group	% of Students Recognizing the Anomaly
Year 7 (n=223)	2%
Year 9 (n=188)	9%
Year 11 (n=122)	19%

Source: Adapted from Lubben and Millar (1996)

Table 1.5. Students preferred representations of data

Students' preferred representation of the data	% of students who preferred this representation of the data when the anomaly was identified			% of students who preferred this representation of the data when the anomaly was not identified		
	Year 7 (n=223)	Year 9 (n=188)	Year 11 (n=122)	Year 7 (n=223)	Year 9 (n=188)	Year 11 (n=122)
Represent the data set using the result of the first, the middle, or the last trial (rationale based in sequence of data collection).	-	-	-	9%	2%	2%
The Mode	-	-	-	44%	32%	21%
The Mean	-	-	-	27%	47%	51%
The Median	-	-	-	3%	2%	4%
One of the extremes (maximum or minimum value in the data set)	-	-	-	8%	3%	2%
The Mode (anomaly excluded)	0%	1%	2%	-	-	-
The Mean (anomaly included)	0%	2%	3%	-	-	-
The Median (anomaly excluded)	1%	7%	13%	-	-	-

Source: Adapted from Lubben and Millar (1996)

In a similar vein of research, Allie et al. (1998) found that undergraduate physics students were split regarding the handling of an anomalous data point. About a third included the data point because all data must be included in the mean. Another third chose to exclude it if it was out of an acceptable range.

For most of the students in the Allie et al. (1998) study, the mean (with or without inclusion of the anomaly) was the preferred way of representing a data set. Conceptions of the mean as they pertain to the accurate representation of a data set were also examined in the Lubben and Millar (1996) study. In this phase of the study, students inspected two data sets with the same mean but different variance and were asked to choose the set that seemed more reliable. Students' responses are outlined in Table 1.6.

Table 1.6. Students' views of the mean and variance of a data set

Student Responses	Year 7 (n=223)	Year 9 (n=188)	Year 11 (n=122)
The data set with smaller variance is more trustworthy.	15%	37%	39%
The data set with larger variance is more trustworthy.	5%	7%	9%
Both data sets are equally trustworthy because they have the same average.	42%	38%	40%

Source: Adapted from Lubben and Millar (1996)

These results are very similar to those of Allie et al. (1998) who found that approximately half of the undergraduate physics students in their study considered variance in their evaluation of the reliability of data. In addition, approximately half of these students viewed data sets with similar means (but considerably different variances) as equally reliable. One interpretation of these findings is that the students' association of "trustworthiness" with the mean is an indication of an inappropriate notion of representativeness. That is, the preference of the mean over the variance as a measure of reliability likely illustrates a conception of the mean that inappropriately associates it with a typical value of the data set. Other researchers have observed student difficulties with the notion of the mean as a representation of multiple measurements (Garfield & Ahlgren, 1988; Mokros & Russell, 1992).

Many students' tendency to neglect variance as a measure of reliability can lead to other problems regarding data interpretation. For example, students who wish to justifiably compare the means of two or more data sets must also examine the variance of each data set. An analysis of the variances allows the student to more confidently attribute observed differences in the means to either differences made to some independent variable value or to random variation in the data. Such an examination of variances might help students distinguish random experimental error from small but reliable differences in data (Schauble, 1996).

The research literature is suggesting that both children and young adults hold simplistic views of the measurement reliability of scientific evidence. For example, findings in the literature indicate that children and adults often hold naïve notions of: the rationale for repeated trials, appropriate treatment of anomalies, and the role of variance in establishing reliability. The protocols used in this study aimed to extend the existing research by examining the teacher-participants' thinking about these and related issues.

K-16 Students' Conceptions of Experimental Validity

Early in this discussion, the practice of repeating measurements (under relatively constant experimental conditions) was described as a means of establishing the reliability of a data set. The number of measurements taken in an experiment is also connected to the range of and interval between the measurement values in a data set. For example, when students are investigating the relationship(s) between continuous variables (i.e., variables that can be represented by any integral or non-integral numeric value), they must often modify the independent variable to produce values of the dependent variable that are separated by an appropriate interval and distributed over an adequate range. Otherwise, the nature of the relationship(s) between the continuous variables may not be readily interpretable (Gott & Duggan, 1996). Recent research suggests that a large percentage of middle and high school students do not address issues such as the range of and interval between dependent variable values when allowed to design their own data collection strategies (Gott & Duggan, 1995; Strang, 1990). Since issues of data range and interval relate closely to the extent to which experimental results can be generalized, they are essentially external validity issues (Wiersma, 1995).

Another important issue regarding the validity of experimental procedures is the extent to which observed changes in a dependent variable value can be attributed to the purposeful modification of some independent variable value. Establishing the causal link between changes in independent and dependent variable values is related to the internal validity of an experiment (Wiersma, 1995) and requires that extraneous variables be controlled. Research has uncovered students' difficulties with the notion of controlled experimentation. Inhelder and Piaget (1958), for instance, viewed the control of variables as a general index of intellectual development. According to their theory, the concept of variable control should develop by the age of 15 (formal operations stage). Subsequent research has shown, however, that many adults do not possess appropriate conceptions of controlled experimentation (e.g., Renner & Lawson, 1973; Wollman & Lawson, 1977) and children much younger than 15 can distinguish between controlled and uncontrolled experimentation (e.g., Case, 1974; Shayer & Adey, 1981; Stone & Day, 1978; Wollman, 1977).

The lack of consistency apparent in these research findings suggests that conceptions of controlled experimentation are likely influenced by factors other than age or intellectual development. The influence of subject-matter knowledge (of the variables pertinent to the investigation in question) has been demonstrated in studies examining K-12 students' conceptions of controlled experimentation (e.g., Duggan, Johnson & Gott, 1996; Levine & Linn, 1977; Linn, Clement, & Pulos, 1983; Linn & Swiney, 1981; Schauble, Klopfer, & Raghavan 1991). Its influence on conceptions of controlled experimentation that was observed in these studies suggests that one might be less likely to design or conduct a controlled experiment when the relationship(s) between the variables are relatively complex. Since the depth of one's subject matter knowledge tends to vary from context to context, it follows that the likelihood of students controlling all pertinent variables in an experiment is also context specific (Lawson, 1985).

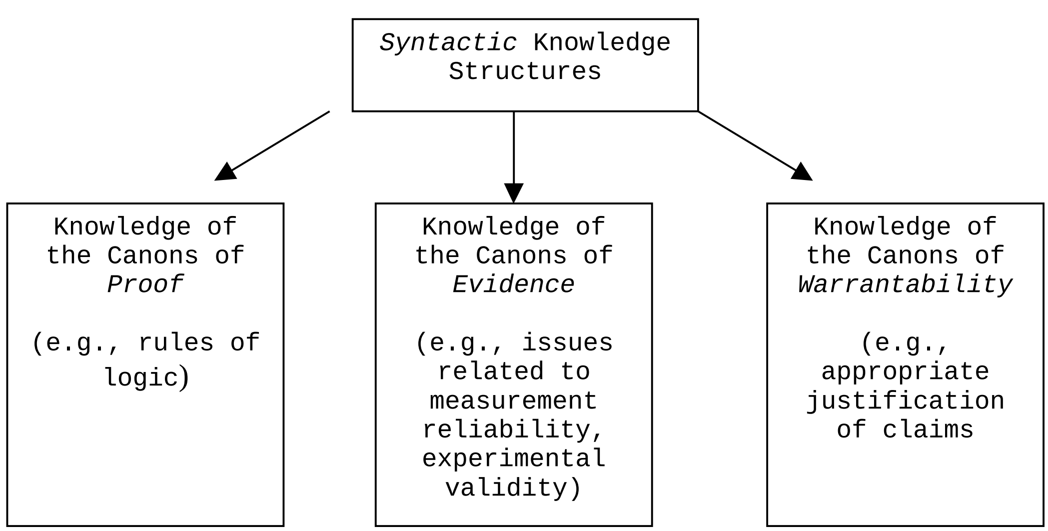
The findings of the research reviewed in this section suggest that K-16 students often hold limited understandings of the experimental validity of scientific evidence. Specifically, misconceptions related to controlled experimentation and to appropriate manipulation of independent variables have been identified in the research literature. In response to these findings, the protocols used in this study targeted these and related conceptions.

Conceptions of the Measurement Reliability and Experimental Validity of Scientific Evidence: A Necessary Dimension of Teachers' Subject Matter Knowledge

Shulman (1986) and colleagues (Grossman, Wilson and Shulman, 1989) identified several categories of teacher knowledge that inform teachers' instructional practices -- subject matter knowledge, general pedagogical knowledge, and pedagogical content knowledge. Subject matter knowledge refers to knowledge of the subject or discipline. General pedagogical knowledge includes teachers' knowledge of teaching, learners and learning and consists of general instructional approaches and strategies for managing the classroom environment. Pedagogical content knowledge (PCK) includes "the ways of representing and formulating the subject that make it comprehensible to others," and "an understanding of what makes the learning of specific topics easy or difficult" (Shulman, 1986, p.9).

Numerous scholars have suggested that teaching for understanding requires teachers to possess a robust understanding of subject matter (e.g., Ball, 1988; Borko & Putnam, 1996; Carlsen, 1991; Grossman, 1990; Magnusson, Krajcik, & Borko, 1999). Some have extended this notion by distinguishing between different types of essential subject matter knowledge (Schwab, 1964, 1978; Shulman, 1986; Smith & Neale, 1989). For example, Schwab (1964) suggested that subject matter knowledge structures include knowledge of the concepts, principles, and theories of the discipline, as well as knowledge of the canons of evidence, proof, and warrantability that guide inquiry in a discipline. Schwab referred to these as *substantive* and *syntactic* knowledge structures respectively. Conceptions of the measurement reliability and experimental validity of scientific evidence are related to the canons of evidence described by Schwab and, therefore, constitute an important dimension of a teacher's syntactic subject matter knowledge (see Figure 1.1).

Figure 1.1. Conceptions of Measurement Reliability and Experimental Validity as a Dimension of Syntactic Knowledge



The importance of conceptions of measurement reliability and experimental validity as a dimension of syntactic subject matter knowledge has not stimulated an adequate line of pertinent teacher education research. Exceptions to this trend include the work of Jungwirth (1985, 1987, 1990) and colleagues (Jungwirth & Dreyfus 1990, 1992). In the Jungwirth (1985) study, prospective and practicing science teachers (primarily life science) responded to hypothetical scenarios that described experiments, data, and in some cases, the conclusions drawn from the data. Some of the scenarios Jungwirth used contained conclusions based on a single observation while others contained conclusions that were based upon insignificant differences in data. These scenarios were grounded in biological as well as “everyday” (non-curricular) contexts.

Jungwirth (1985) asked the science teachers to respond to these scenarios by selecting among several different “opinions” of the students’ experiment (see sample items in Figure 1.2). Table 1.7 illustrates how a sample of 39 South African science teachers (29 in-service and 10 student teachers) responded to the items shown in Figure 1.2.

Figure 1.2. Sample multiple-choice items: Jungwirth (1985)

1. 150 members of a sports club prepared for a marathon. Group A (50 members) took part in 20 training sessions. Group B (50 members) took part in 15 training sessions. Group C (50 members) took part in 10 training sessions.

48 members of group A successfully completed the marathon. 46 members of group B successfully completed the marathon. 44 members of group C successfully completed the marathon.

What is your opinion?

(a) The results were to be expected, since it is well known that in sports those who train more succeed better.
(b) The difference between the two groups is too small to allow conclusions.
(c) In this case the results show clearly that an increase in training results in an increase in achievement.
(d) I don't agree with any of these choices

2. A grade 8 class performed the following experiment: They grew one bean plant at 10 degrees Celsius and another at 30 degrees Celsius. All other conditions (soil, water, light, etc.) were the same. After several weeks, the plant grown at 30 degrees Celsius was almost twice as tall as the other one and much better developed.

What is your opinion?

(a) The experiment shows that a temperature of 30 degrees Celsius is much better for beans than one of 10 degrees Celsius.
(b) It is well known that warmth is needed for plant development, so the results could be expected.
(c) There are many different kinds of beans, some like higher and some like lower temperatures, and this explains the results.
(d) I don't agree with any of these choices.

Source: Adapted from Jungwirth (1985)

Table 1.7. Science teachers' responses to Items 1 and 2

Respondent	% Selecting Option B in Item 1	% Selecting Option D in Item 2
Science Teachers (n = 39)	27%	6%

Source: Adapted from Jungwirth (1985)

The findings of Jungwirth (1985) indicate only a small percentage of the science teachers were concerned that only one bean plant was tested at each temperature. In addition, only a small percentage of the teachers recognized that the differences across groups in the number of sports club members finishing the marathon were quite small and could have resulted from either random variation or factors other than training time. Jungwirth and colleagues made similar observations in their subsequent studies of prospective and practicing science teachers (Jungwirth, 1987, 1990; Jungwirth & Dreyfus 1990, 1992).

Although the work of Jungwirth and colleagues informs science teacher educators as to the nature of science teachers' conceptions of appropriate sampling techniques and statistical significance, their studies were conducted primarily with life science teachers and did not include scenarios that described physical science-based data or conclusions. Scholars such as McPeck (1981) suggest that what counts as "good" evidence might very well differ from one science domain to the next. Therefore, it is also worthwhile to investigate the conceptions of appropriate sampling techniques and statistical significance that physical science teachers use when evaluating physical science-based data or conclusions.

It is also important to note that conceptions of sampling and statistical significance, although important, represent only two of a much larger set of conceptions that relate to the measurement reliability and experimental validity of scientific evidence. Additional conceptions (e.g., experimentation, treatment of outliers) have been described in the literature as being influential to the evaluation of scientific evidence (see previous two sections). The nature of these conceptions, as held by practicing and prospective science teachers, has gone largely unexplored in recent research. Therefore, additional conceptions (see Appendix A) also warrant attention in teacher education research.

Comments on the Scope and Limitations of This Study

For the purposes of this study, a framework was constructed to describe the conceptions most closely related to the scientific evidence (see Appendix A). The existing frameworks of Gott and Duggan (1996), Varelas (1997), Lubben and Millar (1996), Jungwirth (1987), and AAAS (2001) were instrumental to the development of this study's framework.

The *Atlas of Science Literacy* (AAAS, 2001) was particularly influential to the development of this study's framework. This document describes a set of conceptions similar to those examined in this study. However, in the

Atlas of Science Literacy, one must look across several different concept maps to find all of the concepts described in this study’s framework. In fact, one must examine the concept maps entitled: *Evidence and Reasoning in Inquiry* (page 17), *Scientific Investigations* (p. 19), *Averages and Comparisons* (p. 123), and *Statistical Reasoning* (p. 127) to find all of the conceptions of scientific evidence targeted in this study. Although the classification (of conceptions) scheme employed in the *Atlas of Science Literacy* is appropriate in a broad sense, I believe there are conceptions described in this document that are potentially related to scientific evidence but are not classified as such and could be more explicitly connected to the concept map entitled *Evidence and Reasoning in Inquiry*. Connections that are more explicit might help K-16 teachers use the maps in the *Atlas of Science Literacy* to design science curricula that address, in a comprehensive way, issues related to scientific evidence.

In synthesizing the existing frameworks that illustrate conceptions of scientific evidence, I came to appreciate the challenge of thoroughly describing the breadth of conceptions that pertain to scientific evidence.

Therefore, in an effort to streamline the target concepts of this study, it was necessary to limit this investigation to only those conceptions of scientific evidence related to measurement reliability and experimental validity issues.

The focus of the study was further refined by the nature of the implications that I intended to make. Specifically, since this study aimed to draw implications specific to secondary school physics teacher education, I focused on only those conceptions of measurement reliability and experimental validity that are most applicable to the secondary school context.

Notes to the Reader

This dissertation is organized around three main chapters (chapter 3-5) each representing a sub-study of the larger research project. Each of these chapters is written in manuscript style, meaning the chapter is designed to and read like an article in a research journal. The three manuscripts (chapters 3-5) included in this dissertation were customized to address the following limitations in the existing research literature: chapter 3 examines conceptions of scientific evidence that have received little attention in previous research; chapter 4 extends the work of Jungwirth and colleagues to physics teachers and to the evaluation of physical science-based data and conclusions; and chapter 5 describes the usefulness of the research protocols used in this study and suggests other protocols that might also provide insight into conceptions of scientific evidence.

In the remaining chapters, the phrase, “conceptions of the measurement reliability and experimental validity of secondary school-based evidence” will sometimes be referred to simply as “conceptions of the reliability and validity of scientific evidence” or “conceptions of scientific evidence.” Use of the simpler language was deemed necessary to preserve clarity and simplicity in sentence structure.

The design and procedures employed in this study of physics teachers’ conceptions of scientific evidence are described in the next chapter. Then, in chapters 3, 4 and 5, the various findings are discussed. The dissertation ends with a summary, implications and recommendations in Chapter 6.

CHAPTER 2

LINKING THE PURPOSES OF THIS STUDY WITH APPROPRIATE RESEARCH DESIGN

A Qualitative Case Study Design

The principal purpose of this study is to address the following question: *what* is the nature of prospective and practicing physics teachers' conceptions of the measurement reliability and experimental validity of scientific evidence? Addressing "what" questions typically warrants intensive description and interpretation (Creswell, 1998). Consequently, this study utilized qualitative case study research methods in an effort to collect in-depth and comprehensive information about the research participants' conceptions of scientific evidence.

Merriam (1988) broadly defined a case study as "an examination of a specific phenomenon" (p. 9) "to bring about understanding that in turn can affect and perhaps even improve practice" (p. 32). Stake (1995) defined a case as having specific boundaries in terms of phenomenon, time, and place. For the purposes of this study, the boundaries were a set of three physics teachers' collective conceptions of the measurement reliability and experimental validity of secondary-school based scientific evidence.

Since the case investigated in this research could be described more extensively if broken down into sub-units, specific conceptions of scientific evidence (e.g., the rationale for repeating trials) were treated as sub-units for analysis and described separately. This analysis strategy was based upon the recommendations of Yin (1994), who suggested that, in some instances, the decomposition of a case into a series of sub-units or "sub-cases" might facilitate a more thorough analysis. Specifically, Yin stated that the analysis of sub-units "can often add significant opportunities for extensive analysis, enhancing the insights into the single case" (p. 44). Because this approach was taken, much of the data analysis took place within the context of a single conception. A more thorough description of the data analysis techniques used in this study is provided later in this chapter (see section titled *Data Analysis*).

Sampling Techniques

In the words of Patton (1990), the emphasis of the research design used in this study was to facilitate “illumination, understanding, and extrapolation rather than causal determination, prediction, and generalization” (p. 424). Further, Patton (1980) suggested that a “purposeful” sampling technique (like the one used in this study) might be a potentially effective way to gather diverse data. Other qualitative research scholars such as Merriam (1988) and Glaser and Strauss (1967) recommend strategies similar to purposeful sampling for the selection of case study participants. According to Glaser and Strauss, the case study participants should be chosen to maximize or minimize “both the differences and the similarities of data” (p. 55). To maximize potential diversity in the data, participants were recruited voluntarily from several points in the career span. Betty was in the early stages of a teacher education program, Kurt was a first year teacher, and Nina was an 11-year veteran. They all taught or were being prepared to teach secondary school physics.

The sampling technique was not intended to support the generalization of findings across the career span but was instead employed to allow a more thorough exploration of each sub-case through a diversity of data. Each research question was explored using the collective data of all three research participants and, therefore, can be thought of as a *collective case study* (Stake, 1995).

The Participants

Betty, although early in a teacher education program, had already completed seven calculus-based physics courses. Two of these courses covered topics in mechanics while two others focused on topics in electricity.

Betty did not have an undergraduate minor field or an advanced academic degree. She had not conducted an original research project in physics nor had she taken any courses in research design or statistics.

Although Betty had not yet begun her student teaching experience, she held an undergraduate teaching assistantship as a physics lab instructor for a semester before this study. Betty indicated that her duties as a part of this position were focused primarily on grading homework and laboratory reports. Betty reported that she did not frequently interact with the students about their experimental procedures, data, or conclusions.

Kurt, having completed the same teacher education program that Betty was enrolled in, completed a similar number and credits in physics as Betty. He too had taken two courses in mechanics and two in electricity. Kurt did not complete an undergraduate minor field or hold an advanced degree. He had not conducted original research in physics nor had he taken a course in research design or statistics. Kurt was in his first year of high school physics teaching.

Nina, certified also in biology and chemistry, was in the 11th year of her career and was responsible for teaching all three subjects at her school. Nina had completed six undergraduate physics courses. She too had taken at least two courses in both mechanics and electricity. Nina had not completed original research in physics nor did she possess an advanced degree of any kind. She too had not completed a course in research design or statistics.

Informed Consent

Informed consent was obtained at the time participants agreed to be a part of the study. The purpose of the study was explained to the participants and informed consent forms were distributed at that time. After providing each participant adequate opportunity to seek clarification and ask questions, the participants were asked to indicate consent by submitting the Informed Consent Form (see Appendix B).

Data Collection: A Detailed Look

The objective of the following section is to describe the two primary research protocols that were designed for this study. The *Think-Aloud Experimental Design Interview* and the *Analysis of Classroom Passages* generated the primary data for this study. Each type of protocol contained two sub-protocols so that participant's conceptions could be explored in two different physics contexts. The description of each type of protocol includes the rationale for its use, the details of its development, and the way it was used to collect data. This section concludes with a summary of the data sources available in this study.

Interview Protocols: The "Think-Aloud" Experimental Design Interviews

Rationale

Interviews were used in this study because they have been recommended by researchers as a potentially effective way to probe the details of the participants' thinking and have allowed researchers to provide a rich description of their participants' conceptions (Posner & Gertzog, 1982). Specifically, interviews have been effective in revealing students' conceptions of scientific evidence. In fact, many of the research studies described in Chapter 1 successfully utilized interview techniques (e.g., Allie, Buffler, Kaunda, Campbell, & Lubben, 1998; Lubben & Millar, 1996, Varelas, 1997).

A growing number of science education researchers have also recommended the use of task-based interview protocols (e.g., Driver, Squires, Rushworth, & Wood-Robinson, 1994; Schoenfeld, 1985, White & Gunstone, 1992). Specifically, White and Gunstone (1992) suggested, “an even more powerful probe of understanding involves the student in carrying out a task with objects, while answering questions about reasons for each action” (p. 76). This is consistent with other researchers’ support of the use of in-depth interviews that surround the completion of a task (e.g., Lubben & Millar, 1994; McDermott, 1990b; Osborne & Freyberg, 1985; Posner & Gertzog, 1982).

A protocol that is consistent with these recommendations is the *think-aloud* protocol (for a full description see Schoenfeld, 1985). In the think-aloud method, research participants are given one or more tasks and are asked to describe what they are thinking as they complete each task. Schoenfeld found the method effective in probing students’ thinking while solving geometry problems. However, he suggested the method be augmented with other protocols to better probe understanding. Schoenfeld also found the method flexible, claiming that it might be easily adapted to examine students’ thinking in other disciplines.

In addition to eliciting verbal responses during task-based interview protocols, some science education researchers have asked their participants to construct written artifacts of their thinking. For example, in a study of fifth-grade students’ conceptions of experimentation, Baxter, Shavelson, Goldman, and Pine (1992) supplemented their analysis of student performance (in designing and conducting an experiment) by asking the students’ to construct an experiment notebook. In this notebook, the students described their experimental procedure as well as the rationale behind their conclusions (also see use of notebooks in Ruiz-Primo, Baxter, & Shavelson, 1993). After scoring both the performance task and the notebook, Baxter, Shavelson, Goldman, and Pine found that the scores were highly correlated and therefore suggested the possibility that written artifacts might also be a reliable source of data for examining conceptions of experimentation.

Informed by these recommendations for task-based protocols and the collection of written artifacts, the data collection approach used in this study included a semi-structured, task-based adaptation of the Schoenfeld (1985) “think-aloud” protocol referred to as a *Think-Aloud Experimental Design Interview* as well as occasional requests for handwritten responses. A detailed description of both the think-aloud interview protocol and the nature of the hand-written artifacts are provided in the following section.

Description of the Think-Aloud Protocol

The tasks that the participants worked through during these interviews included the design of several experiments. One experiment was to investigate the relationship between a wire's length and its resistance. Another experiment was to determine the relationship between the weight of a wooden block and the minimum applied force necessary to pull the block up an inclined plane. A third was to investigate the relationship between the angle of an inclined plane and the minimum applied force necessary to pull a block up the plane. An outline of the electricity-based version of the *Think-Aloud Experimental Design Interview* is provided in Appendix C as an exemplar.

Each experiment was designed for use with the equipment I provided (e.g., power supply, ammeters, spring balances, etc.). In addition to articulating their thinking as much as possible, the participants were asked to construct a handwritten outline of the experiment they were describing. This provided access to both interview transcripts and written artifacts.

The questions asked in the two *Think-Aloud Experimental Design Interviews* (both inclined plane experiments were discussed in the same interview) focused on the participants' conceptions of the reliability and validity of scientific evidence and their knowledge of pertinent science subject matter (i.e., wire resistance and inclined plane mechanics).

Prompts one and two of each *Think-Aloud Experimental Design Interview* were intended to explore pertinent subject matter knowledge by asking the participants to identify all the variables that affect a given dependent variable (e.g., applied force, wire resistance) and to describe the nature of each relationship. Prompts three through seven in each interview were designed to examine specific conceptions of the measurement reliability and experimental validity of the scientific evidence by asking each participant to describe portions of their design rationale.

Using the participant's proposed experimental design, I devised hypothetical scenarios that described student-generated data and conclusions. The participants were then asked, in prompts eight through twelve, to react to the data and conclusions described in these scenarios. This approach was preferred over asking the participants to collect data in an actual experiment and interviewing them about their data. The rationale behind this preference was based upon several concerns. There was concern that any "real data" that may be collected in an actual experiment may render the focus questions inapplicable. For example, had a participant collected data that did not include an anomaly, there would not have been an opportunity to authentically examine that participant's conceptions of the treatment of anomalies. In addition, it was suspected that the actual collection of data by the participants (enough data to support a response to some of the prompts) would have required the participants to donate an unreasonable amount of time to the study.

The Interviews were audio taped and video taped. At each interview, laboratory equipment was present for use as "props" to support some of the interview prompts. There were several purposes of videotaping the interviews: capturing any participant manipulation of the laboratory equipment, providing a video record of the participants' notes and experiment outlines (sometimes written on a nearby dry erase marker board), and providing a backup audio data source. Both the audio and videotapes were transcribed for analysis.

Interview Protocols: The Analysis of Classroom Passages

Rationale

As noted in Chapter 1, only a few different research protocols have been field tested for studying conceptions of scientific evidence. One protocol that has received some attention involves responding to hypothetical classroom scenarios. Nott and Wellington (1995a, 1995b) used hypothetical classroom scenarios referred to as “critical incidents” to investigate science teachers’ conceptions of the reliability of scientific data. A typical scenario described student investigations that did not produce desirable data. The teachers reviewed the scenarios and responded to questions such as “what would you do” or “what should you do?”

Ehud Jungwirth asked science teachers to respond to hypothetical classroom scenarios as well. His scenarios were constructed in two different formats: a multiple-choice protocol (1985, 1987) called the *TAMI Test* (see sample item in Chapter 1- Figure 1.2) and a more open-ended protocol that required extended responses to the hypothetical scenarios (1990). Over the past twenty years, the use of extended response protocols in teacher education research has garnered increasing support as a reliable measure of selected critical thinking skills (e.g., Jungwirth, 1990; Kitchener & King, 1981; Solano-Flores, Jovanovic, Shavelson, & Bachman, 1999).

Consequently, this study incorporated hypothetical scenarios for use in conjunction with the *Think-Aloud Experimental Design Interviews*. Specifically, a survey referred to as an *Analysis of Classroom Passages* [one in a mechanics context (inclined plane) and one in an electricity context (resistance in a wire)] was developed especially for use in this study (see Appendices D and E).

Description of the Classroom Passages Protocol

Completion of an *Analysis of Classroom Passages* required each participant to respond to two prompts (one and two) intended to examine pertinent subject matter knowledge and to eight hypothetical scenarios or “passages” (prompts *3* through *10*) aimed at investigating specific conceptions of the reliability and validity of scientific evidence (see Table 2.1).

Table 2.1. Target conceptions for prompts 3 through 10 in the *Analysis of Classroom Passages* protocols

Prompt Number	**Target Conception**
3	Instrument Choice (Range and Precision)
4	Controlled Experimentation
5	Manipulation of Independent Variables (Range and Interval)
6	Variance as a Measure of Reliability
7	Recognition and Treatment of Outliers
8	Measurement Validity
9	Generalization of Conclusions
10	Significant Differences in Data

Prompts one and two were designed to assess the participants’ pertinent subject matter knowledge by asking them to identify all the variables affecting a given dependent variable (i.e., applied force, wire resistance) and to describe the nature of each relationship. In prompts three through ten, a hypothetical scenario that described a student-designed physics experiment was presented and each participant was asked to place him or herself into the role of a secondary school physics teacher. As the physics teacher, each participant was to respond to the student-collected data and/or conclusions.

In each of the passages, either the students' data collection strategies were in some way flawed or the students drew inappropriate conclusions based upon their data or both. This protocol required that the participant provide typed responses to the students' data and/or conclusions within an electronic copy of the protocol. This was requested to avoid the situation where illegible hand-written responses could not be interpreted; it also circumvented the need for response transcription and facilitated the entry of data into the qualitative data analysis software used in this study.

For the purposes of clarification or elaboration, the participants were occasionally questioned further about their word-processed responses. These questions usually originated from my efforts to confirm the accuracy of the field notes. Data checking interactions such as these took place at the conclusion of every meeting with the participants (regardless of protocol). The data checking interactions were both audio taped and video taped as well as summarized in my field notes.

Using a strategy similar to that used in the *Think-Aloud Experimental Design Interviews*, the participants were provided with a set of laboratory equipment to possibly enrich their responses to the *Analysis of Classroom Passages* protocols. The participants were instructed, when applicable, to assume that the experiments described in the scenarios were conducted using the provided equipment. Again, the rationale for video taping the participants included capturing the participants' manipulation (if any) of the laboratory equipment while responding to the scenarios as well as providing a backup audio data source.

Data Collection Procedure

In the *Think-Aloud Experimental Design Interviews*, I was interested in observing whether the participants would explicitly and spontaneously address controlled experimentation in their experimental design. Consequently, in this study, the *Think-Aloud Experimental Design Interviews* were conducted first, to reduce the likelihood of a "practice effect" (Kubiszyn & Borich, 1990) that may have resulted from the participants' consideration of other research prompts.

It was suspected that if the participants were to complete *the Analysis of Classroom Passages* protocol first, they may actually learn from the prompts in that protocol.

Specifically, a participant who wrestled with the issue of controlled experimentation in the *Analysis of Classroom Passages* protocol (inclined plane context) might become more sensitive to the need for controlled experimentation. This increased sensitivity could have biased the participants' experimental design in the *Think-Aloud Experimental Design Interview* (inclined plane context). That is, the likelihood that each participant would have spontaneously (without prompting) designed a controlled inclined plane experiment would have been unduly increased. To minimize possible practice effects, both of the *Think-Aloud Experimental Design Interviews* (one in each context) were conducted first and all participants completed the series of four protocols in the same order. In addition, several weeks were allowed to elapse in between meetings with each participant.

Each of the three physics teachers donated approximately four hours to this study. These four hours were donated during four, one-hour meetings where a single protocol was completed. All but a few minutes of these meetings were audio taped and video taped resulting, after transcription, in approximately 30 pages of transcripts per participant. In addition to the word-processed transcripts, I had access to other word-processed artifacts (electronic files from the *Analysis of Classroom Passages*), hand-written artifacts such as experimental design outlines, and approximately 8 pages of field notes per participant (see Table 2.2).

Table 2.2. The types and approximate quantity of data collected from the participants in this study

Protocol	Field Notes (pages)	Transcripts (pages)	Electronic Artifacts	Hand-written Artifacts
Think-Aloud Experimental Design Interview-Inclined Plane	2	10	No	Yes
Think-Aloud Experimental Design Interview-The Resistance of a Wire	2	10	No	Yes
Analysis of Classroom Passages-The Inclined Plane	2	6	Yes	No
Analysis of Classroom Passages-The Resistance of a Wire	2	6	Yes	No

Data Analysis

The participants' word-processed responses to the *Analysis of Classroom Passages* protocol, as well as their transcribed responses to requests for clarification and/or elaboration were analyzed in a similar fashion to the transcribed responses from the *Think-Aloud Experimental Design Interviews*. Each word-processed transcript was reviewed and subdivided into units of information consisting of one or more sentences within the participants' responses. Lincoln and Guba (1985) recommended that qualitative analysis of data begin with this unitizing or open-coding process. The decision as to how much information to include in a unit was influenced by my desire for each unit to adequately represent the participants' reasoning. Consequently, it was occasionally necessary for a unit to represent two or more conceptions. In retrospect, this flexibility in the unitization of data likely facilitated the description of concept integration that became a primary theme of this study. Lincoln and Guba also suggested that a unit of information should be pertinent to the research questions, as well as interpretable in and of itself. There were times when this was possible with one-statement units but many times entire paragraphs were included in a single unit.

The word-processed units of information were imported into the qualitative analysis software program QSR NUDIST^{tm} (Qualitative Solutions and Research Pty Ltd., 1993). Once part of a QSR NUDIST^{tm} project file, the text units were selected (see Figure 2.1) and coded by participant, by context, by protocol, and by emerging category (see Figure 2.2). In QSR NUDIST^{tm}, these categories are referred to as "free nodes."

Data that were coded by participant, context, and protocol were assigned a three-number code. For example, a response from Betty (participant 1) in the electricity context (context 2) to the *Think-Aloud Experimental Design Interview* (protocol 1) was coded (1 2 1). Similarly, a response from Betty in the electricity context to the *Analysis of Classroom Passages* protocol was coded (1 2 2).

Figure 2.1. A selected text unit within QSR NUDIST^{tm}

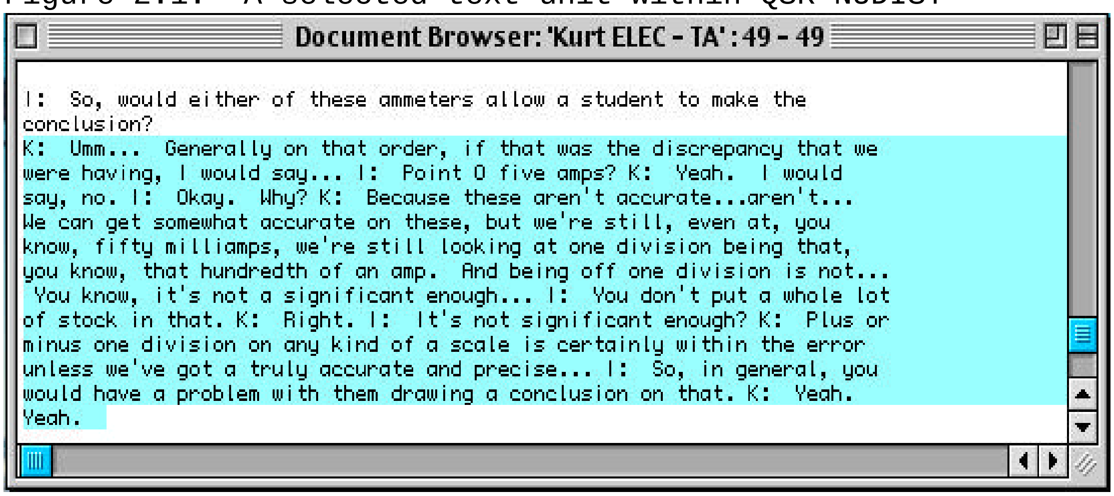
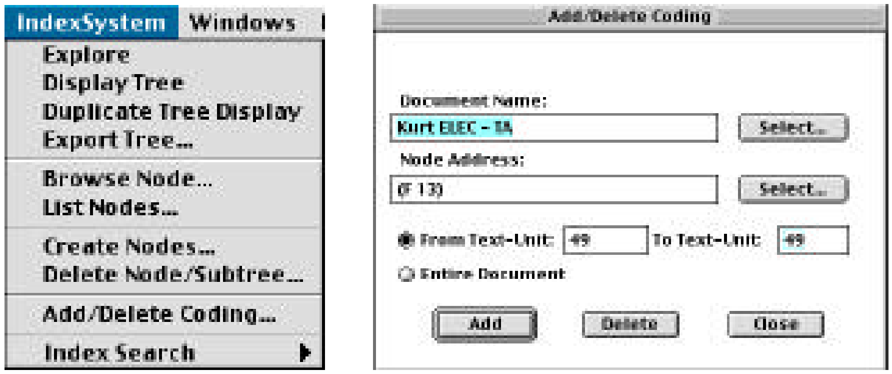


Figure 2.2. Adding coding to selected text within QSR NUDIST™

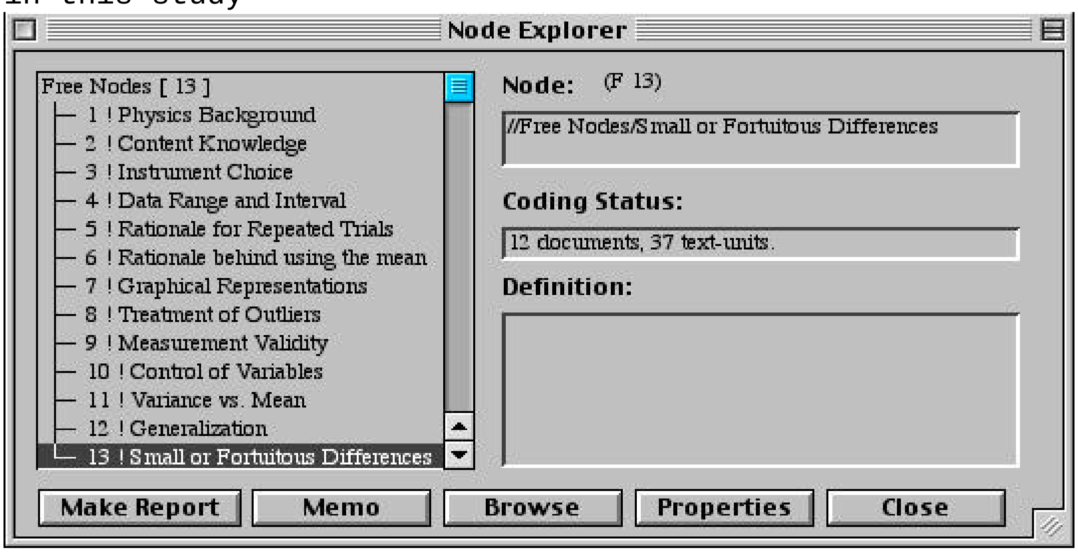


The coded units of information that comprised each free node served as the basis for the final set of data categories. I attempted to maintain a tight connection between these emerging categories and the purposes of the study.

To further develop each category, recurring regularities in the data were identified. At times, as a new unit was placed in a category, the categories were refined. This required a constant comparison between new and existing units of information within a given category (see Creswell, 1998). This method has been referred to in the qualitative research literature as "the constant comparative method" (e.g., Bogdan & Biklen, 1998; Glaser & Strauss, 1967; Merriam, 1988; Strauss & Corbin, 1994).

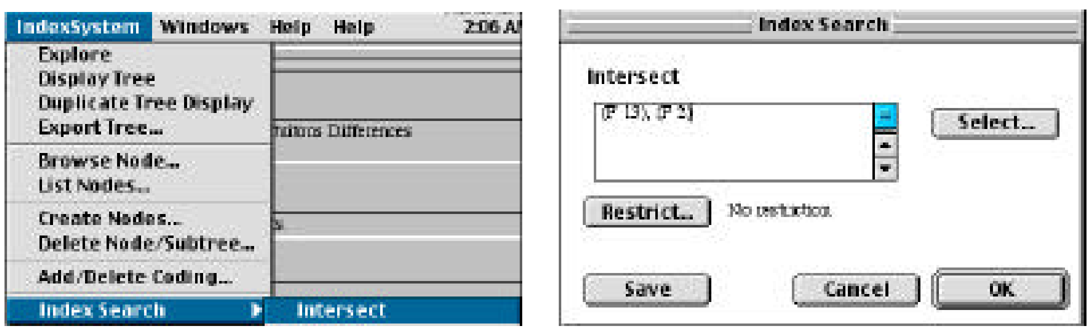
The decision that the final set of data categories was adequately comprehensive was based upon two observations: a reasonably small number of unassigned units and minimal ambiguity in any classification decisions. That is, by the conclusion of the data sorting and category refinement phase, most new data units could be easily classified. Holsti (1969) suggested that qualitative researchers consider similar issues when evaluating the comprehensiveness of existing data categories. The final set of data categories (or free nodes) and their respective data units supported the description of the sub-cases examined in this study (see Figure 2.3).

Figure 2.3. The thirteen Free Nodes or Sub-Cases analyzed in this study



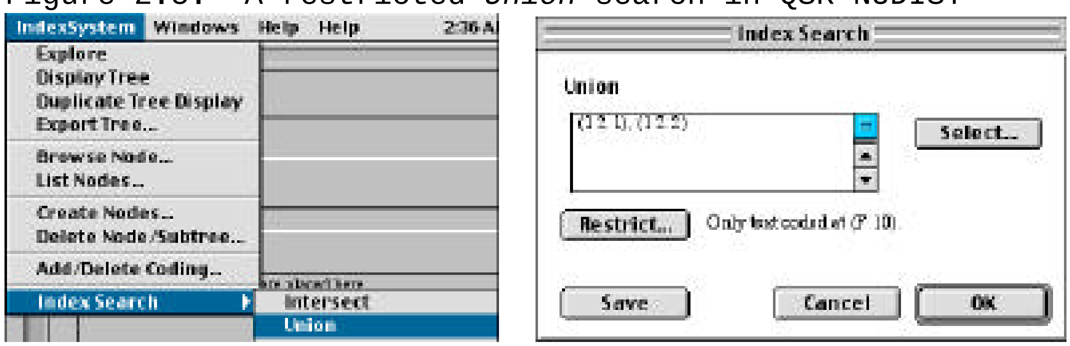
QSR NUDIST^{tm} was also used to automate data searches aimed at exploring and pursuing emerging themes. For example, the degree of integration that occurred between subject-specific science concepts (coded F 2) and conceptions related to the statistical significance of differences in data (coded F 13) was investigated by using a text search called “*Intersect*.” Using the *intersect* function (see Figure 2.4), all of the data units that pertained to *both* the statistical significance of differences in data as well as subject-specific science concepts could be gathered to form their own category or node.

Figure 2.4. Use of the *Intersect* search function to investigate concept integration



Another major theme that was elaborated in this study was the influence of different protocol formats on the nature of the participants’ responses (see chapter 5 for full details). The use of another QSR NUDIST[tm] text search called “*Union*” was most helpful in looking at the participants’ responses across protocols. For example, one cross-protocol investigation examined the nature of Betty’s conceptions of controlled experimentation within the electricity context. The *Union* function facilitated this investigation by gathering all the responses that Betty made in the *Analysis of Classroom Passages* (electricity context – coded 1 2 2) and the *Think-Aloud Experimental Design Interview* (electricity context – coded 1 2 1). The results of the *Union* search were then restricted to only those text units that were *also* associated with conceptions of controlled experimentation (coded F 10). Figure 2.5 illustrates how this restricted *Union* search was conducted.

Figure 2.5. A restricted *Union* search in QSR NUDIST^{tm}



Evaluating the Methods Used in This Study with Established Quality Criteria

The question of whether or not the results of this study are truly trustworthy can be addressed by comparing the research methods with the quality criteria for naturalistic inquiry described by Erlandson, Harris, Skipper, and Allen (1993). Erlanson, et al. suggested that the trustworthiness of the findings of a qualitative research study could be expressed in terms of their *credibility, transferability, dependability,* and *confirmability*. The following sub-sections will discuss the extent to which the findings of this study met these four conditions.

Credibility

Erlandson et al. (1993) defined credibility as the "compatibility of the constructed realities that exist in the minds of the inquiry's respondents with those that are attributed to them" (p. 30). Extending the work of Guba and Lincoln (e.g., Guba & Lincoln, 1981, 1989) Erlandson et al. suggested several strategies for establishing this compatibility. These strategies include: *persistent observation, prolonged engagement, triangulation, referential adequacy materials, and peer debriefing*. The following sections describe how each of these strategies was used in this study.

Persistent Observation and Prolonged Engagement

Erlandson et al. (1993) suggested that thorough description in case study research could be obtained only by “consistently pursuing interpretations in different ways in conjunction with a process of constant and tentative analysis” (p. 30). In this study, data were collected from each participant during four separate meetings. At least two weeks elapsed between meetings. In the time between each meeting, I examined any written or typed artifacts as well as my field notes. Consequently, an analysis of the participants’ responses occurred both at the conclusion of as well as constantly throughout the data collection phase of this study. These preliminary analyses resulted in initial interpretations that guided the construction of questions to be asked at subsequent meetings with the participants.

Because two or more weeks elapsed between meetings with each participant, the data collection phase lasted between two and three months. The prolonged nature of the data collection phase decreased the likelihood that I would collect data that were influenced by the participants’ involvement in some unusual or temporary circumstance.

Triangulation

Triangulation, according to Creswell (1998) is the process of "corroborating evidence from different sources to shed light on a theme or perspective" (p. 202). Two types of triangulation were examined in this study*: data source triangulation* and *methodological triangulation.*

Data source triangulation involves the corroboration of evidence collected from the same data source (usually a person) but in different contexts or under different conditions. Methodological triangulation usually requires a comparison of data across the different research protocols used in a particular study.

For the purposes of this study, the data sources were the three participants. Therefore, data source triangulation in this study aimed to examine each participant's ability to critically evaluate scientific evidence across different physical science contexts. For example, each participant's responses to the *Analysis of Classroom Passages* protocol (inclined plane context) were compared to his or her responses to the *Analysis of Classroom Passages* protocol (electricity context).

Specifically, I focused on comparisons across responses to parallel items.

In general, the participants' responses to parallel items were consistent in indicating a certain degree of reliability within specific types of research protocol. For example, when comparisons were made between the participants' responses to item four on the electricity-based *Analysis of Classroom Passages* protocol (Appendix D) and their responses to item four on the inclined plane-based *Analysis of Classroom Passages* protocol (Appendix E), a relatively high level of consistency was observed. For example, Betty's responses to these items are included in the following excerpts:

<u>Betty</u>

Analysis of Classroom Passages (Electricity):

> B: This really can't be judged if you are looking at how length affects resistance because you are using different materials and like I said before the different materials may have different properties in themselves that affect the resistance so, in order to test for length you probably want to use the same material at a few different lengths rather than varying the material and the lengths.

Analysis of Classroom Passages (Inclined Plane):

> B: They need to control the angle that the incline is at. You can't just change that from one trial to the next.

A high degree of response consistency was observed across most sets of parallel items and was often confirmed by my field notes.

The methodological triangulation involved cross-protocol comparisons. That is, each participant's response to a given prompt was compared to his or her response to one or more additional prompts that differed in format but shared the same target conception. As described in detail in chapter 5 of this dissertation, these cross-protocol comparisons uncovered several inconsistencies, which were most visible in prompts intended to examine conceptions of independent variable manipulation and controlled experimentation. One interpretation of the inconsistencies is that the protocol used in this study had limited *construct validity* (Gronlund, 1998). A more optimistic interpretation might conclude that the slight differences between the natures of each protocol helped me make certain distinctions between tasks related to the evaluation of scientific evidence. More specifically, evaluative tasks where conceptions of evidence were generally used in conjunction with subject-specific science concepts could be distinguished from those tasks that did not necessarily require such integration of knowledge.

Referential Adequacy Materials

Because data must be interpreted in terms of its context, it is necessary for researchers to collect a variety of artifacts to give a holistic view of the context (Erlandson, Harris, Skipper, & Allen, 1993, p. 31). In this study, audio taped interviews were also video taped to capture the participants’ manipulation of materials. In addition, the participants provided typed and hand-written responses to selected research prompts. These artifacts, taken collectively, provide a detailed picture of the research context.

Peer Debriefing

Erlandson et al. (1993) recommended that researchers occasionally step out of the context of their research and seek guidance and feedback from a qualified colleague.

Throughout the course of this study, I met with the thesis advisor to review research foci, analyses, and implications. Often, these meetings resulted in a more refined and focused research plan. In addition, the thesis advisor's interpretation of the data enriched the analyses by confirming my initial perceptions or, at times, by encouraging alternative explanations.

Transferability

The purpose of this study did not include the generalization of findings to some larger population of prospective or practicing physics teachers. However, this is not to say that the findings have no relevance to other contexts. Provided that other contexts share one or more characteristics with the context of this study, a degree of transferability may exist (Erlandson, Harris, Skipper, & Allen, 1993, p. 32).

The reader evaluates the transferability of this study by examining the intricacies of its context. Therefore, I attempted to describe the context of the study in great detail to help the reader with this decision. Erlandson et al. (1993) refer to the detailed documentation of context as *thick description* and recommend it as an important transferability criterion.

To assist the reader in evaluating transferability, I described in detail several aspects of this study. The purpose, research questions, and design of this study were elaborated. I described the participants with respect to their academic backgrounds, teaching experience, and involvement in original physics research. I also described the types of protocols used in this study and provided these protocols (for further reference) in the appendices. Further, my description of the findings of this study included references to detailed interview excerpts. The use of lengthy interview excerpts was required to establish the context in which the participants’ statements were made.

Dependability and Confirmability

The detailed description of the research methods and findings of this study also provides the reader with a comprehensive *dependability audit* (Erlandson, Harris, Skipper, & Allen, 1993, p. 34). A dependability audit is the documentation of research methods and findings necessary for the reader to judge whether a replication of the study with similar respondents in similar contexts might yield similar findings. However, since invariance across the findings of qualitative case studies is improbable, the description of the research methods and findings (dependability audit) should be detailed enough to allow the reader to attribute any variation in the findings to particular sources (e.g., different insights, error).

The description of the research findings of this study included both interview excerpts that allow the reader to connect my interpretations directly to the data and the explicit reasoning used to assemble my interpretations and implications. This documentation provides a *confirmability audit* (Erlandson, Harris, Skipper, & Allen, 1993) that allows the reader to determine if the conclusions, interpretations, and recommendations: can be traced to their sources, and are indeed supported by the research methods.

Content Validity

Although not explicitly recommended by Erlandson et al. (1993), another criterion commonly used to evaluate the trustworthiness of the methods or findings of a study is the *content validity* of the instrumentation or research prompts. Gronlund (1998) suggested that the issue of content validity asks the question: do these research prompts adequately represent the content domain in question? In this study, the content domain that the research prompts aimed to represent consisted of conceptions of the measurement reliability and experimental validity of scientific evidence as applicable to secondary school science.

Part of the rationale for including open-ended prompts in the research protocols in this study was to promote diversity in the (reliability and validity related) issues mentioned by the participants. It was hoped that the diverse data might include mention of issues that had not been anticipated or previously identified in the review of pertinent literature. In general, the participants did not mention reliability or validity issues that were completely unanticipated, which suggested that the breadth of concepts examined provided a degree of content validity to the study.

At times, however, subtle aspects of some of the more common conceptions of scientific evidence were uncovered and these subtle aspects were integrated with other concepts in some unique ways. One example of this unique use of concepts was observed in both Kurt’s analysis of the significance of differences in data (see full description in chapter 4) and in his rationale for repeating trials. It was expected that Kurt would mention something related to the notion of “spread” between a set of measurement values and indeed Kurt articulated similar ideas. However, Kurt’s evaluation of the variance of a set of data was apparently influenced by his conceptions of instrument precision. This included his ideas regarding both appropriate estimation techniques (between two demarcations on a instrument scale) and measurement tolerance. Further, the notion that a measurement might best be accompanied by a tolerance estimate (an indication of possible instrument related error) was mentioned only by Kurt and has not been examined with any regularity in pertinent research. Therefore, the results of the content validity analysis done at the conclusion of this study suggest that future research regarding conceptions of the reliability and validity of scientific evidence might explicitly examine conceptions of measurement tolerance and estimation and/or how these concepts influence one’s rationale for repeating trials or one’s evaluation of differences in data.

Summary

The purpose of this qualitative case study was to describe one prospective and two practicing physics teachers’ conceptions of the measurement reliability and experimental validity of scientific evidence. The data collected to inform the case study was collected from both task-based and scenario-based protocols and was organized by individual conception before being analyzed. The software program QSR NUDIST^{tm} supported the data analysis. The following chapters contain illustrations of the main findings of this study.

CHAPTER 3

INTEGRATING UNDERSTANDINGS: A CASE STUDY OF SECONDARY SCHOOL PHYSICS TEACHERS' EVALUATION OF SCIENTIFIC EVIDENCE

For the past several decades, science educators have recommended scientific literacy be improved among this nation's citizenry (e.g., Bauer, 1992; Dewey, 1933; Herron, 1971; Kyle, 1980, Yager, 1991). These recommendations identified the ability to critically evaluate scientific evidence or knowledge claims as an essential component of scientific literacy. In more recent science education reform documents, one's ability to critically evaluate scientific evidence has been specifically linked to appropriate conceptions of scientific inquiry (NRC, 1996) and of the nature of science (AAAS, 1993).

Some in the science education community have suggested that the concepts one uses when thinking critically about scientific evidence be viewed as a type of subject matter knowledge (e.g., Gott & Duggan, 1996; Lubben & Millar, 1996). Specifically, the ability to critically evaluate evidence can be, in part, supported by a distinct set of conceptions regarding scientific evidence. The rationale behind the present study was strongly influenced by this view.

Recent research has contributed to a growing understanding of students' conceptions of scientific evidence (e.g., Carey, Evans, Honda, Jay, & Unger, 1989; Gott & Duggan, 1995). Identification of some alternative conceptions of scientific evidence has been central to the findings of many of these studies. Difficulties were described for students in elementary school (Varelas, 1997), in middle and high school (Foulds, Gott, & Feasey, 1992), and in undergraduate-level physics courses (Allie, Buffler, Kaunda, Campbell, & Lubben, 1998). Some of these studies examined aspects of students' conceptions of measurement reliability such as the need for repeated trials (Schauble, 1996), the treatment of outliers (Chinn & Brewer, 1993), and the variance of data sets (Allie, Buffler, Kaunda, Campbell, & Lubben, 1998). Other studies focused on aspects of students' conceptions of experimental validity such as controlled experimentation (Schauble, Klopfer, & Raghavan, 1991) and appropriate data collection strategies (Strang, 1990).

In light of the findings of these and similar studies, the *National Science Education Standards* (NRC, 1996) and

Science for All Americans (AAAS, 1989) suggested methods for helping students develop the ability to critically evaluate scientific evidence. For instance, the vision of science learning described in the *National Science Education Standards* included engaging students in “the presentation of scientific evidence, reasoned argument, and explanation” (p. 50). To this end, this document suggested the importance of the teacher’s role in facilitating classroom discourse regarding scientific evidence. It describes the teacher’s role in this environment as one who guides decisions as to which “ideas to follow, ideas to question, information to provide, and connections to make” (p.36). Similarly, *Science for All Americans* (AAAS, 1989) suggested that students need guidance in “collecting, sorting, and analyzing evidence, and in building arguments based on it” (p.201).

These suggestions raise an important question for science teacher educators. What types of understandings should science teachers possess in order to provide such guidance? Scholars and researchers in teacher education have suggested for some time that teaching for understanding requires a thorough understanding of the subject matter (e.g., Ball, 1988; Borko & Putnam, 1996;

Carlsen, 1991; Grossman, 1990; Schwab, 1978; Shulman, 1986; Smith & Neale, 1989). It follows that science teachers must first possess appropriate conceptions of scientific evidence themselves before they can provide the support necessary for their students to develop similar conceptions. Yet, research regarding prospective and practicing teachers’ conceptions of scientific evidence is not highly visible in the existing literature. Consequently, neither pre-service nor in-service science teacher education has been properly informed by carefully designed research related to teachers’ conceptions of scientific evidence.

Purpose of Paper

This paper is part of a larger study that endeavored to describe and interpret the nature of prospective and practicing physics teachers’ conceptions of scientific evidence. The descriptive case study research reported in this paper is one of the sub-studies in the larger research project and focuses on conceptions related to the measurement reliability and experimental validity of scientific evidence (see examples in Table 3.1). This paper addresses the following research questions:

- What types of issues related to the measurement reliability and experimental validity of scientific evidence do the participant-teachers think about when designing experiments?

- When presented with hypothetical scenarios that describe unsound experimental procedures or poorly supported conclusions (or both), what concerns will the participant-teachers raise?

For the purposes of this study, the nature of each participant's conceptions of scientific evidence was extrapolated from the nature of his or her experimental designs and evaluative responses to student-generated scientific evidence or conclusions.

Research Methods

The case under study in this research was secondary school physics teachers' conceptions of scientific evidence. This case was informed by data from multiple participants and can be thought of as a collective case study (Stake, 1995). Since the case being described could be explored more extensively if broken into sub-units, specific conceptions of scientific evidence (e.g., the rationale for repeating trials) were treated as sub-units for analysis and compared across participants.

In this study, participants were selected because they varied in the duration of their physics teaching experience. One participant was recruited from each of the following points in their careers: early in the teacher education program (Betty), during the first year of teaching (Kurt), and after 11 years of teaching experience (Nina). Differences in conceptions of scientific evidence were expected across this span because it was assumed that experience in struggling with student-generated data and conclusions would promote the development of certain conceptions of scientific evidence. Therefore, the differences in teaching experience should result in data that would allow the authors to construct a more thorough description of the case.

Each participant responded to two, semi-structured *Think-Aloud Experimental Design Interviews* (for a full description of the think-aloud method see Schoenfeld, 1985). In the think-aloud method, the participants were given one or more tasks and were asked to describe what they were thinking as they completed each task. The tasks completed by the participants in this study included designing several experiments. One experiment consisted of investigating the relationship between a wire’s length and its resistance. Another was to determine the relationship between the minimum applied force necessary to pull a wooden block up an inclined plane and the weight of the block. A third was to investigate the relationship between the minimum applied force necessary to pull a block up an inclined plane and the angle of the inclined plane surface. These experiments were to be conducted using equipment provided by the authors.

The other protocol used in this research, the *Analyses of Classroom Passages* (see Appendices D and E), required the participants to respond both orally and in writing to a series of hypothetical classroom scenarios that were developed especially for this study. These hypothetical scenarios described student-designed experiments and, when appropriate, corresponding student-generated conclusions. As with the think-aloud interviews, these scenarios dealt with both electricity and inclined plane contexts. The rationale for this protocol was based in part upon the suggestions of previous researchers who found that responses to hypothetical scenarios or passages (like those in this protocol) were potentially reliable measures of selected critical thinking skills (e.g., Jungwirth, 1990; Jungwirth & Dreyfus, 1975; Kitchener & King, 1981).

Findings and Discussion

An analysis of the data collected in this study suggested that each participant integrated subject-specific science concepts with selected conceptions of scientific evidence when evaluating data and/or claims. That is, each participant used subject-specific science concepts in conjunction with selected conceptions of scientific evidence when evaluating that evidence. In addition, the authors observed that some conceptions of scientific evidence (e.g., measurement validity, statistical significance) tended to be integrated with subject-specific science concepts more extensively than other conceptions of scientific evidence (see Table 3.1). These finding are consistent with the results of other research that examined K-16 students' conceptions of scientific evidence (e.g., Foulds, Gott, & Feasey, 1992; Gott & Duggan, 1995; Linn, Clement, & Pulos, 1983; Schauble, Klopfer, & Raghavan 1991).

Upon recognizing the emergence of these general themes, the ensuing cross-participant analysis of specific conceptions of scientific evidence focused on two goals: to describe how the participants' subject-specific science concepts interacted with their conceptions of scientific evidence and describe how certain conceptions of scientific evidence seemed to be more extensively integrated with subject-specific science concepts than others. Table 3.1 presents a summary of the relationships found between conceptions of scientific evidence and subject-specific science concepts.

Table 3.1. Conceptions of scientific evidence: Integration with subject-specific science concepts

More Integrated	**Less Integrated**
· Controlled Experimentation · Generalization of Conclusions · Measurement Validity · Statistical Significances of Differences in Data · Recognition and Treatment of Outliers · Instrument Choice (scale and precision · Manipulation of Independent Variables	· Control of Variables · Rationale for Repeated Trials · Reliability of Data

In the following section, selected cross-participant analyses are provided as examples of differing degrees of integration with subject-specific science concepts.

Extensive Integration of Subject-Specific Science Concepts - Example 1: Controlled Experimentation

In the *Think-Aloud Experimental Design Interview* (electricity context), the participants first identified the factors they viewed as influential to the resistance of a wire. Only one of the participants identified as many as three of the four variables (i.e., resistivity, length, cross-sectional area, and temperature) that influence the resistance of a segment of conducting wire (see Table 3.2). Then, the participants designed an experiment that could be conducted to investigate the relationship between the length of a wire and the wire's resistance. The authors instructed each participant to design the experiment to yield data upon which sound conclusions could be based. The participants, while thinking out loud about the task, constructed a hand-written outline of their preferred experimental design.

After each participant described his or her experimental design, the authors asked the participant to describe why the experiment was a fair test of the desired relationship. Portions of the participants' responses are provided in the following interview excerpts to illustrate their reasoning.

Betty:

> B: To test thirty centimeters for copper and forty centimeters for nickel (wire), you're not getting any kind of comparisons. So, the thing that you would want to do is do like, if you're not gonna do anything else at all, do thirty, forty, fifty, sixty centimeters in just copper. You need to just make sure that the length is the only thing that you are varying when you're doing the test.
> (*Analysis of Classroom Passages – The Resistance of a Wire*)

Kurt:

> I: So, why is it important to hold those things constant?
> K: So you can get some sort of comparison…if we enter in different materials, or different thicknesses we're gonna have differing results from that and that will affect our resistance. So we want to limit the number of external factors from our experiment.
> (*Think-Aloud Experimental Design Interview – The Resistance of a Wire)*

Nina:

> I: So why is it important to keep these factors constant?
> N: So that you can attribute any changes in resistance to changes in length.
> (*Think-Aloud Experimental Design Interview – The Resistance of a Wire)*

These responses indicate a rationale for designing controlled experiments that focuses on students' ability to make meaningful comparisons between measurements. They also suggest a conception of scientific evidence that is based upon the need for obtaining interpretable data.

After examining the participants' hand-written outlines and analyzing their spoken comments, the authors observed that the participants' planned to control only the variables they believed would affect the dependent variable (see Table 3.2).

Table 3.2. Variables identified and later controlled in the *Think-Aloud Experimental Design Interviews*

Think-Aloud Interview Context	Variables identified as influential to Dependent Variable (Minimum Applied Force/Resistance)	Variables explicitly controlled in the experimental design outline
The Inclined Plane Betty	Weight of Block, Friction Forces, Angle of Inclination	Weight, Frictional Forces
Kurt	Weight of Block, Friction Forces, Angle of Inclination	Weight of Block, Friction Forces, Angle of Inclination
Nina	Weight of Block, Friction Forces, Angle of Inclination	Weight of Block, Friction Forces, Angle of Inclination
Wire Resistance Betty	Length, Radius	Length
Kurt	Length, Material, Temperature	Length, Material
Nina	Length, Voltage, Current	Voltage, Current

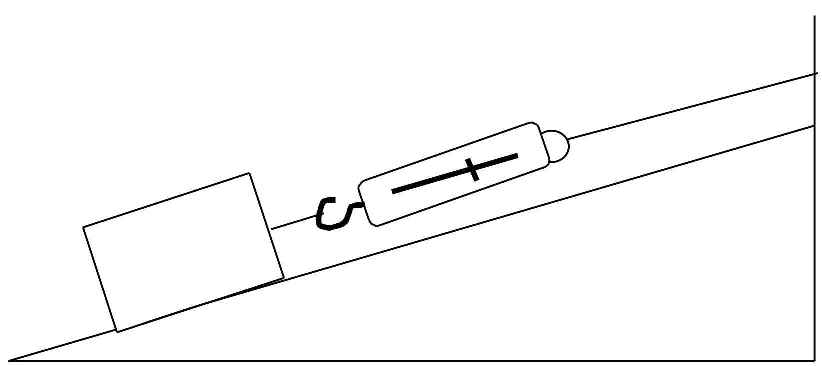
Of course, not all variables initially identified as such were later controlled, but all of the variables that each participant controlled in their proposed experiment were among those that he or she identified earlier as being influential to the dependent variable. The influence of science-specific subject matter knowledge on conceptions of experimental design was also noted in other studies (e.g., Duggan, Johnson, & Gott, 1996; Lawson, 1985; Levine & Linn, 1977; Linn & Swiney, 1981).

Extensive Integration of Subject-Specific Science Concepts - Example 2: Measurement Validity

In some of the hypothetical scenarios presented in the *Analysis of Classroom Passages* protocol, the participants were presented with inappropriate conclusions based upon misuse of instruments (see Figure 3.1).

Figure 3.1. Research prompt: Inclined Plane Classroom Passage

In one presentation, a student group thoroughly described their experimental procedure. The students mentioned that the minimum applied force necessary to pull the block up the incline could be measured by attaching (using string) the end of the spring scale to a post at the top of the incline (see figure below). The students argued that measuring the applied force this way would account for the block's weight, the slope of the incline, and any frictional forces.



If at all, how would your respond to this group?

The participants tended to offer similar responses to this prompt. That is, each participant was concerned that using the spring balance in this way might not be appropriate. The following interview excerpts include the participants' response to the students' argument:

Betty:

> B: It was a good attempt but it doesn't really take into account the frictional forces in the way that you need to because the block is not moving. It does take into account the constant slope of the incline and the weight of the block.
> *(Analysis of Classroom Passages – The Inclined Plane)*

Kurt:

> K: I would discuss with the group the fact that we are investigating the kinetic friction and not the effect of static friction. Their setup does not allow us to investigate the kinetic friction, which comes from the block moving on the incline.
> *(Analysis of Classroom Passages – The Inclined Plane)*

Nina:

> N: Kinetic frictional forces could not be measured because the block is not moving.
> *(Analysis of Classroom Passages – The Inclined Plane)*

These responses indicate that the participants shared a common understanding that the block must be moving with respect to the surface of the inclined plane for the spring balance to measure kinetic frictional forces.

As the participants designed experiments during the *Think-Aloud Experimental Design Interviews*, the authors paid close attention to how each participant planned to use the available equipment in his or her proposed experiment. Kurt, for instance, specified that the spring balance be attached to the block and that the block be pulled up the incline at a constant velocity. He explained:

> K: Well, if you add in acceleration instead of just sliding it along at a constant rate, the balance will also measure the extra force going into the acceleration.

Kurt’s statement illustrates how he was able to integrate his knowledge of spring balances with his understanding of inclined plane mechanics to justify elements of his preferred experimental design.

Extensive Integration of Subject-Specific Science Concepts - Example 3: The Generalization of Conclusions

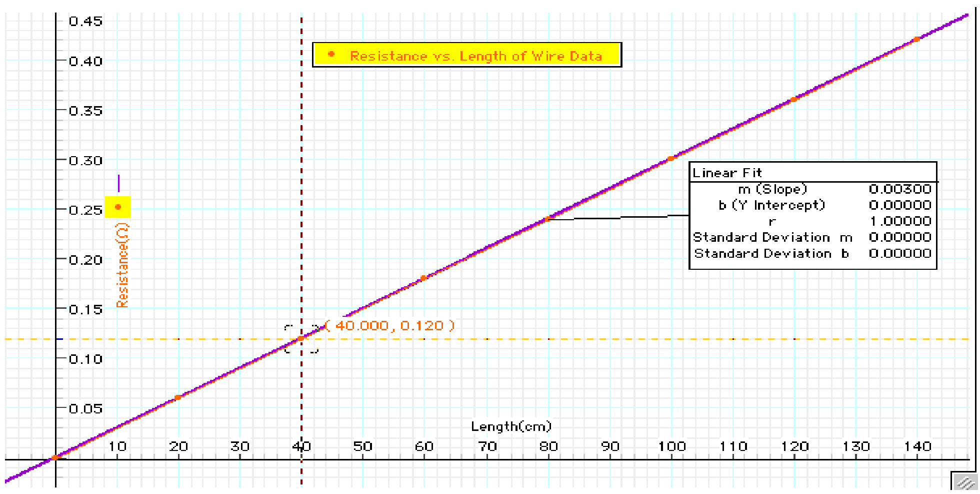
In some of the hypothetical scenarios that the participants considered, students had generalized their conclusions to contexts where they were no longer applicable. Each overgeneralization was based on the students’ interpretation of a graph of their data. Each of the passages, one from each context, is provided in Figure 3.2.

Figure 3.2. Research prompt regarding the Generalization of Conclusions

Referring to the graph below, a student group concluded…

"In sum, the resistance of a wire will increase by 0.003Ω every time its length is increased by 1cm."

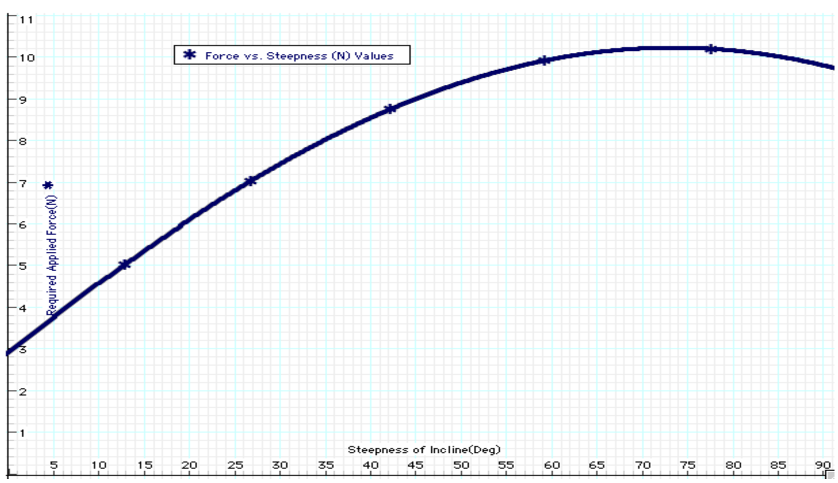
How would you respond to this claim? Why?



Referring to the graph provided, a student group concluded…

"In sum, the minimum applied force necessary to pull an object up an inclined plane will typically reach a maximum when the angle of incline is between 70 and 75 degrees."

How would you respond to this claim? Why?



Of the three participants, only Betty focused largely on the applicability of the students' conclusion. The following interview excerpts contain portions of all three participants' responses to these prompts. Betty's response was to the electricity-based prompt while Kurt and Nina's was to the inclined plane-based prompt.

Betty:

> B: I would agree to this claim because that is what their graph is telling them. The slope of the graph is .003... however they need to clarify their claim because the slope of their graph is .003 but that does not mean the slope of a graph referring to other types of wire will be .003. When they make their claim they need to be specific to their experiment.
> *(Analysis of Classroom Passages – The Resistance of a Wire)*

> B: Like, for the lab I constructed, there were a few different wires that we used and there could be different slopes. Like, for this one, the slope of it is clearly point zero three. So, I would agree with this statement. But they're generalizing by saying the resistance of a wire.
> *(Analysis of Classroom Passages – The Resistance of a Wire)*

Kurt:

> K: I would ask the group about their lab setup and data sampling trying to determine a factor which caused this unusual data. I would also ask them to disregard the data and give their opinion based on understanding without the data. Then I would ask them if they could think of reasons why the data doesn't support our theory and conceptual understanding.
> *(Analysis of Classroom Passages – The Inclined Plane)*

> K: It seems strange that it was dropping off (the required applied force began to decrease at an angle of incline of approximately 75 degrees),

> because that's, you know, that's the portion that I'm concerned with and wondering.
> *(Analysis of Classroom Passages – The Inclined Plane)*
> K: I wouldn't want them to continue on with that thought and would want to then step in to, you know, teaching them that it does continue to increase.
> *(Analysis of Classroom Passages – The Inclined Plane)*

Nina:

> N: Well, it should continue to increase, shouldn't it? That's what I would think. So, I wouldn't agree with this conclusion.
> *(Analysis of Classroom Passages – The Inclined Plane)*

Betty's reaction to the students' conclusions in the electricity-based prompt clearly indicates that she recognized the overgeneralization. Her recognition seemed to be critically related to her understanding that a change in the wire's material and/or diameter would change the slope of the corresponding resistance vs. length graph [the slope identified by the students applies only to a wire where the ratio ρ/A (where ρ = the resistivity of the wire, and A = the cross-sectional area) is approximately equal to $0.003\Omega/cm$].

The students' conclusion in the inclined plane-based prompt includes an overgeneralization in the sense that the point where the minimum applied force reaches its greatest value is influenced by the coefficient of kinetic friction that exists for the surfaces in contact. Therefore, the maximum applied force that occurred between 70 and 75 degrees would occur only when specific materials (block and inclined plane surface) were used. None of the participants questioned the applicability of the conclusion in the inclined plane-based prompt. Kurt and Nina focused their evaluation of the students' conclusion on their observation that the minimum applied force began to decrease when the angle of inclination for the inclined plane approached 80 degrees. This decrease in minimum applied force was inconsistent with Kurt and Nina's expectations. Both Kurt and Nina indicated that they expected the relationship between minimum applied force and angle of inclination to be linear. Kurt and Nina's pre-occupation with the non-linear nature of the graph probably seemed to draw their attention away from the applicability of the conclusion.

Limited Integration of Subject-Specific Science Concepts: Evaluating the Reliability of Data

Some of the scenarios in the *Analysis of Classroom Passages* described data containing multiple measurements taken under identical experimental conditions. The data sets differed from one another in the amount of variance present among the measurement values (see research prompts in Figure 3.3).

Figure 3.3. Classroom Passages related to Reliability of Data

Group D – Inclined Plane Experiment

Trial	1	2	3	4	5	Avg. Applied Force Required
Mass of Block	500g	500g	500g	500g	500g	
Angle of incline	40˚	40˚	40˚	40˚	40˚	
Applied Force Required	6.5N	7.0N	7.5N	7.0N	7.0N	7N

Trial	1	2	3	4	5	Avg. Applied Force Required
Mass of Block	500g	500g	500g	500g	500g	
Angle of incline	40˚	40˚	40˚	40˚	40˚	
Applied Force Required	10.5N	4.5N	9.5N	5.5N	5.0N	7N

Trial	1	2	3	4	5	Avg. Current Measured
Cross-sectional Area of Wire	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	
Length of Wire	50cm	50cm	50cm	50cm	50cm	
Current	1.9A	2.0A	1.8A	2.2A	2.1A	2.0A

Trial	1	2	3	4	5	Avg. Current Measured
Cross-sectional Area of Wire	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	3.4 X 10^{-6} m^2	
Length of Wire	50cm	50cm	50cm	50cm	50cm	
Current	1.0A	3.0A	0.5A	3.5A	2.0A	2.0A

After reviewing these data sets, each participant was asked to indicate which data set he or she viewed as more reliable. Without exception, the participants mentioned that consistency or, in Nina's case, "continuity" in the data was important for reliability. The participants' responses to this question are included in the following interview excerpts.

Betty:

> I: Why do you think that group D's data (electricity experiment) is more reliable?
> B: The data numbers found fell right around the desired average. I think consistent data is more reliable…
> (Analysis of Classroom Passages – The Resistance of a Wire)
>
> I: So, is group E's data (inclined plane experiment) reliable?
> B: No.
> I: Okay. Why not?
> B: Just because it's not around ten. They (the measurements) are just so different, you know, from five to fifteen, I wouldn't think that would be very reliable. Their data is pretty far away from the average they received. But going between nine and eleven and ten, I would think that would be a little more reliable.
> I: So, you think group E has an unreasonable spread.
> B: I think it is. Yeah.
> (Analysis of Classroom Passages – The Inclined Plane)

Kurt:

> I: Why don’t you think group E’s data (inclined plane investigation) is reliable?
> K: Because the data seems to be polar at two extremes, I don't feel that we have enough consistent data to support any true relationships.
> I: So, what's most important to you when you are thinking about reliability?
> K: Consistency with data. But I always tell them, go with the data that you've got if it's good and solid.
> (Analysis of Classroom Passages – The Inclined Plane)

Nina:

> I: What do you think of Group E's electricity experiment?
> N: I would ask Group E to run the experiment again due to the variations in their measurements of applied force… to see if they get a little bit more continuity, I guess. The values are varied too much in comparison to other group findings.
> (Analysis of Classroom Passages – The Resistance of a Wire)

All the participants used the notion of consistency as a criterion for evaluating the reliability of student-generated scientific data. Lubben and Millar (1996) also found in their study of young adults that consistency was frequently used as a criterion for judging the reliability of data sets.

The participants’ subject-specific science concepts were not consistent across contexts. For example, the participants all demonstrated a more thorough understanding of inclined plane mechanics than of wire resistance. When asked to identify the factors that affect the minimum applied force necessary to pull a block up an inclined plane, each participant identified the weight of the block, the angle of the incline, and the friction between the surfaces as influential variables. With the exception of the angle of inclination, each participant accurately described the nature of the relationship between these variables and the minimum applied force. However, as described previously, only one of the participants identified as many as three of the four variables (i.e., resistivity, length, cross-sectional area, and temperature) that influence the resistance of a segment of conducting wire. As a whole, the participants' understandings of the factors that affect the resistance of a wire were diverse. This diversity, coupled with the consistency of responses to data reliability prompts, suggests that concepts regarding reliability were not necessarily integrated with pertinent subject-specific science concepts.

The relative independence of conceptions of the reliability of data from subject-specific science concepts is in stark contrast with the extensive integration cited in the first three examples. The data presented in the first three examples illustrated how the participants integrated conceptions of controlled experimentation, measurement validity, and the generalization of conclusions with subject-specific science concepts. Betty's responses to prompts containing sweeping generalizations illustrated how a deeper understanding of subject specific science concepts might aid the recognition of an overgeneralization. Her understanding that the slope of a resistance vs. length graph is influenced by the wire material apparently made her more sensitive to conclusions containing related overgeneralizations. That is, she was more sensitive to conclusions that did not specify that the slope of the graph applied only to wires of a specific resistivity and cross-sectional area.

Similarly, Kurt and Nina's recognition of an overgeneralization in the inclined plane-based prompt seemed to be inhibited by the existence of certain subject-specific science misconceptions. Specifically, Kurt and Nina's expectation of linearity in the applied force vs. angle of inclination graph that was included in this prompt distracted them from fully attending to the applicability of the stated conclusions.

Implications For Future Research

This study examined a selected set of conceptions related to the measurement reliability and experimental validity of scientific evidence. For the physics teacher participants in this study, this set of conceptions was useful in describing their thinking about scientific evidence in both an inclined plane and an electricity context. This is not to say, however, that one does not incorporate other evidence-related conceptions when designing experiments or evaluating data. Further exploratory inquiry into the possibility of other conceptions of evidence is needed.

Scholars such as McPeck (1981) have taken the epistemological view that what counts as “good” evidence might very well differ from one science domain to the next. Therefore, it is especially important that carefully constructed studies into the possibility of other conceptions of evidence be conducted in a variety of scientific domains.

In this study, the authors grounded their examination of the participants’ conceptions of scientific evidence in the contexts of the inclined plane and the resistance of a wire. These contexts, though important, constitute only two of many possible physics contexts that could have been used in this research. Future research might investigate the extent to which subject-specific science concepts are integrated when evaluating evidence in other physics contexts or in other secondary school science domains (e.g., biology, chemistry, earth science).

The data from this study indicates that the participants integrated their subject-specific science concepts more often with certain conceptions of scientific evidence (control of variables, generalization of data, experimental validity, conclusions based upon small or fortuitous differences, recognition and treatment of outliers, instrument choice, range and interval of data) than with others (control of variables, rationale for repeated trials, reliability of data). Future research involving other physics contexts or other scientific disciplines might also investigate whether certain conceptions of scientific evidence tend to be integrated with subject-specific science concepts more often than others.

It was also observed that the accuracy of certain subject-specific science concepts held by the participants was influential to their ability to identify the contexts to which scientific conclusions might be applied. Future research might explore more deeply the nature of physics teachers’ subject-specific science concepts and how these relate to the generalization of conclusions in other physics contexts as well as how they might otherwise influence the evaluation of scientific evidence.

(Do not type text in this document beyond here)

CHAPTER 4

EVALUATING STATISTICAL SIGNIFICANCE: A COLLECTIVE CASE STUDY OF SECONDARY SCHOOL PHYSICS TEACHERS

Some researchers in the science education community have suggested that concepts involved in data interpretation (such as those related to variance) have been grossly underrepresented in K-16 science instruction (Driver, Newton, & Osborne, 2000; Duschl, 1990). McDermott (1990a) attributed such underrepresentation in part to the prevalence of confirmatory laboratory courses in many secondary and post-secondary science programs, leaving little time for more conceptual studies. This apparent disregard for the development of data interpretation concepts has been linked to many students' inability to critically evaluate scientific claims (Solomon, 1991).

In response to research findings such as these, the *National Science Education Standards* (NRC, 1996) formulated a vision of science learning where students engage in "weighing the evidence, and examining the logic so as to decide which explanations and models are best" (p. 175). This vision is consistent with the views of scholars such as Latour and Woolgar (1986), who described an accurate picture of the nature of scientific activity as including the weighing of evidence/data and the critical assessment of explanations and knowledge claims.

To this end, *Science for All Americans* (AAAS, 1989) suggested that students need guidance in “collecting, sorting, and analyzing evidence, and in building arguments based on it” (p.201). Engaging students in these ways raises an important question for science teacher educators: what types of understandings should science teachers have in order to provide such guidance?

Teacher education scholars and researchers have suggested that teaching for understanding requires rich and flexible subject matter knowledge (e.g., Borko & Putnam, 1996; Grossman, 1990; Shulman, 1986; Smith & Neale, 1989). Similarly, helping students think critically about scientific evidence likely requires that teachers possess appropriate conceptions related to scientific evidence. Specifically, Gott and Duggan (1996) suggested that one’s ability to critically evaluate scientific evidence might be supported in part, by a distinct set of conceptions pertaining to the reliability and validity of scientific evidence. The research described in this paper was influenced by this viewpoint.

Science teachers' conceptions of the reliability and validity of scientific evidence has received little attention in the research literature; only a small number of studies have examined science teachers' evaluations of scientific evidence and knowledge claims (e.g., Jungwirth, 1985, 1987, 1990; Jungwirth & Dreyfus, 1992; Nott & Wellington, 1995b).

Nevertheless, these studies have been important to the field: in them, prospective and practicing (primarily life science) teachers responded to hypothetical scenarios that described experiments, data, and in some cases, conclusions based upon the data. Some of the scenarios contained conclusions that were based upon a single observation while others contained conclusions that were based upon insignificant differences in data. These scenarios were grounded in biological as well as "everyday" (non-curricular) contexts.

In his study, Jungwirth (1985, 1987) asked science teachers to respond to these scenarios in two different ways. He used a multiple-choice protocol, which required the teachers to select among several different "opinions" of the students' experiment (see sample items in Figure 4.1) and, in a 1990 study, he employed a more open-ended protocol in which science teachers provided extended responses to the hypothetical scenarios.

Figure 4.1. Sample multiple-choice items: Jungwirth, 1985,1987

1. 150 members of a sports club prepared for a marathon. Group A (50 members) took part in 20 training sessions. Group B (50 members) took part in 15 training sessions. Group C (50 members) took part in 10 training sessions.

48 members of group A successfully completed the marathon. 46 members of group B successfully completed the marathon. 44 members of group C successfully completed the marathon.

What is your opinion?

(a) The results were to be expected, since it is well known that in sports those who train more succeed better.
(b) The difference between the two groups is too small to allow conclusions.
(c) In this case the results show clearly that an increase in training results in an increase in achievement.
(d) I don't agree with any of these choices

2. A grade 8 class performed the following experiment: They grew one bean plant at 10 degrees Celsius and another at 30 degrees Celsius. All other conditions (soil, water, light, etc.) were the same. After several weeks, the plant grown at 30 degrees Celsius was almost twice as tall as the other one and much better developed.

What is your opinion?

(a) The experiment shows that a temperature of 30 degrees Celsius is much better for beans than one of 10 degrees Celsius.
(b) It is well known that warmth is needed for plant development, so the results could be expected.
(c) There are many different kinds of beans, some like higher and some like lower temperatures, and this explains the results.
(d) I don't agree with any of these choices.

Source: Adapted from Jungwirth (1985)

Table 4.1 illustrates how a sample of 39 South African science teachers (29 in-service and 10 student teachers) responded to the items above. Jungwirth discovered that a

Table 4.1. Science teachers' responses to Items 1 and 2

Respondent	**% Selecting Option B in Item 1**	**% Selecting Option D in Item 2**
Science Teachers (n = 39)	27%	6%

Source: Adapted from Jungwirth (1985)

very small percentage of the science teachers were concerned that only one bean plant was tested at each temperature. In addition, he found that only a slightly larger percentage of the teachers recognized that the differences across groups in the number of sports club members finishing the marathon were quite small and could have resulted from random variation as well as any number of factors besides training time. The small number of teachers recognizing experimental issues becomes especially troublesome when one considers the recommendations of influential reform documents such as *Benchmarks for Scientific Literacy* (AAAS, 1993). This document suggested that students' understanding of the nature of science include the notion that "when similar investigations give different results, the scientific challenge is to judge whether the differences are trivial or significant" (p. 7). Clearly, the judgement process is complicated.

Regarding this process, Jungwirth (1985) described "a lack of knowledge of certain sets of *simple* rules relating to the acceptability or admissibility of evidence and the permissibility of extrapolation in general, and in scientific methodology in particular" (p. 59). Jungwirth's conclusion regarding a limited knowledge about scientific evidence strongly suggests potential foci for research and development issues in science teacher education.

The work of Jungwirth and others has begun to inform science teacher educators as to the nature of science teachers' conceptions of both appropriate sampling techniques as well as statistical significance. However, these studies were conducted primarily with life science teachers and did not examine these conceptions within other disciplines such as physics. This project, then, will focus on research conducted with physics teachers, which requires the evaluation of significant differences in data in physics contexts.

Purpose of the Paper

The principal aim of this research is to describe the collective conceptions of sampling and statistical significance held by a group of three secondary school physics teachers. The focus on sampling and statistical significance is a sub-study in a larger research project that attempts to explain the nature of physics teachers' conceptions of scientific evidence. Specifically, this study was designed to address the following research questions:

- What types of issues related to the sampling of data do practicing and prospective physics teachers consider when designing experiments?
- When presented with hypothetical scenarios that describe unsound experimental procedures or poorly supported conclusions (or both), what concerns related to the sampling of data or the significance of differences in data will the prospective and practicing physics teachers raise?

Research Methods

Since the principal purpose of this study was to describe secondary school physics teachers' conceptions of the significance of differences in data, descriptive case study research methods were deemed most appropriate (Creswell, 1998; Merriam, 1988). Since this case was informed by data from multiple participants, it can be thought of as a *collective* case study (see Stake, 1995). In this study, participants were selected because they possessed varying amounts of physics teaching experience. One participant was recruited from each of the following points in their careers: early in the teacher education program (Betty), during the first year of teaching (Kurt), and after 11 years of teaching (Nina). Differences were expected among these participants because it was assumed that experience with student-generated data and conclusions based on data might promote the development of certain conceptions of scientific evidence. The authors intended to highlight these differences in an effort to thoroughly describe the case.

The protocol used in this research required the participants to respond to two hypothetical classroom scenarios that were developed especially for this study (see item 10 in Appendices D and E). These scenarios, which were grounded in electricity and inclined plane contexts, described student-designed experiments, and when appropriate, corresponding student-generated conclusions. The authors analyzed the participants’ written responses as well as audio taped discussions held with the participants as they constructed their responses. The rationale for this protocol was based in part on previous research that suggested responses to hypothetical scenarios or passages were potentially reliable measures of selected critical thinking skills (e.g., Jungwirth, 1987; Kitchener & King, 1981).

Findings and Discussion

The overarching theme in the findings of this study was that the participants were not always critical of conclusions drawn upon insufficient differences in data. Specifically, the participants did not always recognize flaws in experimental design or statistical insignificance. This pattern was especially evident in the inclined-plane based scenarios. The following illustrative excerpts were taken from the participants' written and oral responses to item 10 of both Analysis of Classroom Passages surveys (Appendices D and E). In the hypothetical scenarios described in item 10 of each survey, the students based conclusions upon differences in data that might be thought of as insignificant or fortuitous. Specifically, the differences in the data could be attributed to limitations in the sensitivity of the instruments or simply to random variation.

<u>Betty (inclined plane)</u>:

I: How would you respond to that?
B: Umm...I would agree with it. But umm...you would think that if they're sliding it along this, this face, there's gonna be a lot more friction acting on it as it moves along.
I: Just to be clear, they were measuring the force that it took to initiate motion, not to keep it moving.
B: Right.
I: So, they took the reading just before it started to move,
B: Right.
B: Okay. I think I would still stick with that answer.

<u>Kurt (inclined plane)</u>:

I: How would you respond to this group's evidence?
K: I would discuss with them and ask them if they thought that one trial for each was sufficient. I would say that those trials by themselves are not enough to umm...garner, you know, enough information to make that kind of a relationship.

K: Also, I would ask them to look at the average for all three trials and look at the distance from the average that each trial was.

K: I would discuss the relative errors that are in this lab. Are the distances from the average enough to give us relationships or are they most probably resulting from the error in this lab? How accurate are our spring scales, Is it plus or minus, you know, a half? Is it plus or minus point one newtons. To see what kind of a range surrounds that value.

<u>Nina (inclined plane)</u>:

I: So they concluded that the applied force required to initiate motion up the plane depends on the face that it's on. What would you...how would you respond to that?

N: Well, I would agree somewhat, Umm...I would say it does vary somewhat.

Betty (electricity):

I: So, what do you think about that conclusion?
B: Well, I don't think that conclusion is correct because I think the current in a series is the same everywhere.
I: So, if they continued to see this difference, what would you...
B: I don't know. I really... But to my knowledge, I would think the current would be the same everywhere in the series circuit. So...
I: So, if the student says, well, what about this difference (in data)?
B: I'm not sure.

Kurt (electricity):

I: How would you respond to this group's evidence?
K: I would discuss with them the error that we have in taking our data and show them that this discrepancy is most likely coming from this error and not from a change in current. If necessary, I would take the students through an error analysis to see just how accurate our data is. This would include how much of our reading was estimated by the students and other variables that could introduce error to our lab.

K: I Would probably talk them through errors in our...in our ammeters.
I: So, would this ammeter allow a student to make the conclusion (Each participant was shown an ammeter with a range of 0-500mA and scale demarcations every 10mA)?
K: Umm... Generally on that order, if that (the difference in data shown in the table) was the discrepancy that we were having, I would say no.
I: Why?
K: We can get somewhat accurate on these, but we're still looking at one division being that, you know, that hundredth of an amp. And being off one division is not... You know, it's not a significant enough...
I: It's not significant enough?

> K: Plus or minus one division on any kind of a scale is certainly within reasonable error unless we've got a truly accurate and precise...
> I: So, in general, you would have a problem with them drawing a conclusion based upon those differences?
> K: Yeah. Yeah.

Nina (electricity):

> I: How would you respond to this group's evidence?
> N: I would think the values (current) should be the same. That's what I would think because of the series circuit. I would need to do the experiment to agree with my students' findings. I think that the students need a greater difference to prove the hypothesis.

Consistent with the findings of Jungwirth (1985, 1990) the interview excerpts suggested that the participants did not always recognize experimental design flaws or the insignificance of differences in data. Further, the participants’ responses indicated that their evaluation of the significance of differences in data was influenced by their subject-specific science knowledge. That is, the participants’ recognition of the statistical insignificance of the differences in data reported in these scenarios was, at times, inhibited by limitations in their understanding of physics concepts. For example, in the inclined plane context, Betty did not express concern with the students’ conclusion. This observation seemed to follow logically from the remainder of her response, which indicated that she expected the amount of force necessary to initiate motion to vary predictably with the area of the surfaces in contact. Similar limitations in subject matter knowledge of mechanics have been documented in research with students *and* teachers of physics (e.g., Finegold & Gorsky, 1991; Palmer, 1997; Trumper, 1996).

It should be noted, however, that Betty's response in the electricity context suggested that a critical evaluation of the significance of differences in does not depend solely on subject-specific physics knowledge. She correctly expected that measurements of electric current in a series circuit should be similar, and this alerted her to a problem with the students' conclusion. But, her understanding of this property did **not** seem useful in helping her to think about how she would help students reflect upon the differences in the measurement values.

Nina's responses also demonstrated the influence of her subject-specific science knowledge. She did not express concern with the differences in applied force measurements reported in the scenario nor did she express an expectation that these values should be similar. In contrast, her expectation of constant current values in the electricity experiment seemed to focus her attention on the significance (magnitude) of the discrepancy in the current measurements. Along with her concern regarding the magnitude of the difference between electric current measurements, Nina also expressed a concern with the number of trials conducted in the experiment:

> I: How would you respond to this group's evidence?
> N: Additional trials would be needed to support their evidence. One trial just isn't enough to really conclude that statement.
> (Analysis of Classroom Passages – The Resistance of a Wire)

Nina’s concern about limited trials clearly focused on the number of trials taken at each ammeter location. Her apparent focus on repeated trials in this research prompt was consistent with a previous statement in which she mentioned that the need for repeated trials had been “drilled into” her in her undergraduate physics courses.

Kurt also expressed concern with the single trial experimental design. The solution he proposed included not only the incorporation of additional trials but also an examination of the variance within those trials. Kurt’s concern with the conclusions drawn by the students was also based upon issues related to instrument precision. He mentioned the amount of estimation that occurs when the needle of an instrument falls between two subdivisions on a scale and how this estimation influences the precision of the measurements. In addition, Kurt's reference to a measurement being "plus or minus" a certain value suggests that he associates a certain amount of unreliability (tolerance) with each measurement. His response indicates that his evaluation of the significance of a difference in data involves knowledge of the reliability of the instrument being used.

In sum, the collective responses of the participants suggested that several different types of subject-matter knowledge were integrated to evaluate the significance of differences in data. Collectively, the participants accessed their understandings of subject-specific science concepts (e.g., friction), instrumentation (e.g., precision), experimental design (e.g., sampling), and variance (e.g., confidence intervals) while making this type of judgment.

Implications for Future Research

Currently, the literature base lacks adequate breadth to properly inform science teacher education as to practicing and prospective teachers' conceptions of experimental design and statistical significance. In this study, the authors grounded their examination of the participants’ conceptions in the contexts of the inclined plane and electrical circuits. These contexts, though important, constitute only two of many possible physics contexts that could have been used in this research. Future research might augment the findings of this study and those of Jungwirth and his colleagues by investigating science teachers’ evaluation of experimental design and statistical significance in other physics contexts or in other secondary school science domains (e.g., chemistry, earth science).

Future research might also examine some other intriguing issues that emerged in this study. For example, Betty described her undergraduate physics experience as one where the need for repeated trials had not always been “emphasized” as much as it had with Nina. The difference in Nina and Betty’s teaching experience raises important questions about the differences in their respective rationale for repeating trials. It seems possible that in the time since Nina’s teacher preparation program, advances in science-specific educational technology such as the development of sensitive probes and computer-based data collection techniques may have inspired a change in how undergraduate physics investigations are designed. Further, these same technological advances may have initiated a change in how scientific inquiry and the nature of science are represented in undergraduate physics courses.

If such a change has indeed occurred, it is possible that Betty's conceptions of the need for repeated trials reflect the implicit messages being sent by contemporary practices in undergraduate physics education. Future research might investigate whether or not the discrepancy in rationales (for repeating trials) observed across Betty and Nina's responses is in fact widespread across novice and veteran science teachers. Further, future research might investigate the influence of prospective science teachers' perceptions of the reliability of computer-based data collection techniques on their rationale for repeating trials.

CHAPTER 5

EXPLORING SCIENCE TEACHERS' CONCEPTIONS OF EXPERIMENTATION: TOWARD A DIVERSE SET OF RESEARCH PROTOCOL

Some science educators have contended that students' critical thinking skills develop as a result of laboratory work; direct instruction of thinking skills is thus unnecessary (Gott & Duggan, 1996). Consequently, critical thinking skills were not historically emphasized in the design of K-16 science curriculum, and serious attempts to define the specific conceptions that underlie critical thinking have not been undertaken.

Similar evidence that critical thinking skills have been under-emphasized is visible in the nature of the research protocol in recent science teacher education research. Lederman, Wade, and Bell (1998), for example, suggest that most of the established *nature of science* assessments do not emphasize the critical evaluation of experimentation and evidence. In fact, science teachers' conceptions of experimentation have received little attention in the research literature. Specifically, only a small number of studies have examined science teachers' evaluation of scientific procedures and evidence (e.g., Jungwirth, 1985, 1987, 1990; Jungwirth & Dreyfus, 1992; Nott & Wellington, 1995b). Consequently, only a few different research protocols have been field tested for science teacher research.

Nott and Wellington (1995a, 1995b) used hypothetical classroom scenarios referred to as “critical incidents” to examine teachers’ conceptions of the reliability of scientific data. In their research, a typical scenario described a student investigation that did not produce reliable data. After the teachers reviewed the scenarios, Nott and Wellington asked them to respond to questions such as, “As the teacher, what would you do?”

Similarly, Jungwirth and colleagues (e.g., Jungwirth 1985, 1987; Jungwirth & Dreyfus, 1990, 1992) asked the science teachers in their research to respond to hypothetical classroom scenarios. The scenarios were posed to the teachers in two different formats. A multiple-choice protocol, called the *TAMI Test*, required the respondents to select among several different “opinions” of the experiment described in the hypothetical scenario (see sample items in Figure 5.1). In a 1990 study, Jungwirth employed a more open-ended protocol that required the science teachers to provide extended responses to the hypothetical scenarios. Support for the use of scenario-based protocols in teacher education research has increased over the past twenty years due to of a growing body of research suggesting that they can be reliable measures of selected critical thinking skills (e.g., Jungwirth, 1990; Kitchener & King, 1981; Solano-Flores, Jovanovic, Shavelson, & Bachman, 1999).

Figure 5.1. Sample multiple-choice item: TAMI Test (Jungwirth, 1985,1987)

1. 150 members of a sports club prepared for a marathon. Group A (50 members) took part in 20 training sessions. Group B (50 members) took part in 15 training sessions. Group C (50 members) took part in 10 training sessions.

48 members of group A successfully completed the marathon. 46 members of group B successfully completed the marathon. 44 members of group C successfully completed the marathon.

What is your opinion?

(a) The results were to be expected, since it is well known that in sports those who train more succeed better.
(b) The difference between the two groups is too small to allow conclusions.
(c) In this case, the results show clearly that an increase in training results in an increase in achievement.
(d) I don’t agree with any of these choices

2. A grade 8 class performed the following experiment: They grew one bean plant at 10 degrees Celsius and another at 30 degrees Celsius. All other conditions (soil, water, light, etc.) were the same. After several weeks, the plant grown at 30 degrees Celsius was almost twice as tall as the other one and much better developed.

What is your opinion?

(a) The experiment shows that a temperature of 30 degrees Celsius is much better for beans than one of 10 degrees Celsius.
(b) It is well known that warmth is needed for plant development, so the results could be expected.
(c) There are many different kinds of beans, some like higher and some like lower temperatures, and this explains the results.
(d) I don’t agree with any of these choices.

Source: Adapted from Jungwirth (1985)

Another important theme in the recent recommendations of science education researchers is the call for task-based protocols (e.g., Driver, Squires, Rushworth, & Wood-Robinson, 1994; Schoenfeld, 1985, White & Gunstone, 1992). Specifically, White and Gunstone (1992) suggested, “an even more powerful probe of understanding involves the student in carrying out a task with objects, while answering questions about reasons for each action” (p. 76). In this, they are consistent with other researchers who have supported the use of in-depth interviews that surround the completion of a task (e.g., Lubben & Millar, 1994; McDermott, 1990b; Osborne & Freyberg, 1985; Posner & Gertzog, 1982).

One research protocol that addresses some of the task-embedded research recommendations is the *think-aloud protocol* (for a full description see Schoenfeld, 1985). In the think-aloud method, research participants are given one or more tasks and are asked to describe what they are thinking as they complete each task. Schoenfeld (1985) found the method to be generally effective in probing students’ thinking while solving geometry problems, but he suggested that it be augmented with other research procedures to comprehensively probe understandings.

Schoenfeld also found the method to be flexible; he suggested that the protocol might be easily adapted to examine students' thinking in other disciplines.

In addition to eliciting verbal responses during task-based interview protocol, some science education researchers have asked their participants to construct written artifacts of their thinking. For example, in a study of fifth-grade students' conceptions of experimentation, Baxter, Shavelson, Goldman, and Pine (1992) supplemented their analysis of student performance (in designing and conducting an experiment) by asking the students' to compose an experiment "notebook." In this notebook, the students described their experimental procedure (including the variables controlled) as well as the rationale behind their conclusions (also see use of notebooks in Ruiz-Primo, Baxter, & Shavelson, 1993). After scoring both the performance task and the notebook, Baxter, Shavelson, Goldman, and Pine found that the scores were highly correlated, suggesting that written artifacts might also be a reliable source of data for examining conceptions of experimentation.

Some studies in science education have integrated several of the research protocols described previously.

For example, Solano-Flores, Jovanovic, Shavelson, and Bachman (1999), in their study of fifth-grade students' conceptions of experimentation, asked the students to investigate relationships and test hypotheses. Solano-Flores et al. then analyzed the students' performances in selected tasks (conducting experiments), written artifacts (experiment design notebooks), and responses to hypothetical scenarios. The hypothetical scenarios used in their study described fabricated data sets. These fabricated data were used to avoid the situation where invalid or unreliable data from the actual experiments might cloud an investigation of the students' conceptions of data analysis.

The Purpose of this Paper

The research reported in this paper is part of a larger study that had the primary purpose of describing the nature of prospective and practicing physics teachers' conceptions of the measurement reliability and experimental validity of scientific evidence. However, this particular paper focuses primarily on the research question:

- When prospective and practicing physics teachers' responses to parallel research prompts are compared across multiple research protocols, what similarities and differences exist in their conceptions of scientific evidence?

Research Methodology: An Integrated Approach

Building on the recommendations of previous researchers, this case study of three secondary school physics teachers' conceptions of scientific evidence utilized several research protocols. The first involved hypothetical scenarios used to determine each of the teacher's conceptions of controlled experimentation and appropriate manipulation of independent variables. The scenarios used in this study also utilized fabricated data to increase the likelihood that the participants would address targeted conceptions in their responses. The scenarios used in this study, referred to as *Classroom Passages*, dealt with inclined plane mechanics and the electrical resistance of a wire (see sample scenarios in Figure 5.2), two physics concepts commonly taught in schools.

Figure 5.2. Sample scenarios: Controlled Experimentation/Representative Sampling

3. One of the student groups claimed that increasing the length of the wire caused the resistance to decrease. The students supported this claim with following data table.

Length of Wire	**Resistance**
Tungsten - 100cm	**200 ohms**
Aluminum -105cm	**175 ohms**
Copper - 110cm	**165 ohms**
Silver - 115cm	**140 ohms**

What is your evaluation of this claim?

4. Three of the student groups claimed that the slope of the incline affected the necessary applied force. The data tables constructed by these student groups are shown below.

Group A

Slope of Incline	5°	65°	70°	80°	85°
Required Force	4.8N	10.6N	10.5N	10.3N	10.1N

Group B

Slope of Incline	18°	36°	54°	72°	90°
Required Force	6.8N	8.9N	10.2N	10.5N	9.8N

Group C

Slope of Incline	1°	2°	3°	4°	5°
Required Force	4.1N	4.3N	4.4N	4.6N	4.8N

How would you evaluate these claims?

The other research protocol was a semi-structured, task-based adaptation of the Schoenfeld (1985) method, referred to as a *Think-Aloud Experimental Design Interview*. The tasks required the participants to design several experiments: one was to investigate the relationship between a wire’s length and its resistance, another was to investigate the relationship between the weight of a wooden block and the minimum force necessary to pull the block up an inclined plane, and a third was to investigate the relationship between the angle of an inclined plane and the minimum force necessary to pull a block up it. In addition to verbally articulating their thinking as much as possible, the participants were asked to construct a handwritten outline of the experiment they designed.

Findings

Conceptions of Controlled Experimentation

In the *Think-Aloud Experimental Design Interview* (electricity context), the participants were first asked to identify the factors that affect the resistance of a wire. In his or her response, no participant identified more than three of the four influential variables (length, resistivity, cross-sectional area, and temperature). The participants’ responses are illustrated in Table 5.1.

Table 5.1. The participants' ideas regarding the variables that affect the resistance of a wire

Participant	Variables that affect the resistance of a wire
Betty	Length, Radius
Kurt	Length, Material, Temperature
Nina	Length, Voltage, Current

The participants then constructed a hand-written outline of an experiment that could be conducted to investigate the relationship between the length of a wire and the wire's resistance. The experiment was to be conducted using only the equipment provided (e.g., power supply, ammeters, etc.). The participants' experimental designs called for controlling only the variables that were viewed as influential to the dependent variable. This is not to say that all variables identified were later controlled, but all variables that the participants controlled in their proposed experiments were among those identified as influential to the dependent variable (resistance). In short, if the participant did not expect a variable to affect the dependent variable, it was not controlled.

In their responses to the hypothetical scenario (item 3 in Figure 5.2), the participants (Betty, Kurt, and Nina – all pseudonyms) immediately recognized the uncontrolled experiment. The following excerpts illustrate their reaction to the scenario:

Betty:

> B: I think I would first ask them their purpose in changing the material and length.
> (Analysis of Classroom Passages – The Resistance of a Wire)

> B: This really can't be judged if you are looking at how length affects resistance because you are using different materials and like I said before the different materials may have different properties in themselves that affect the resistance so, in order to test for length you probably want to use the same material at a few different lengths rather than varying the material and the lengths.
> (Analysis of Classroom Passages – The Resistance of a Wire)

Kurt:

> K: I would try to help them see the mistake by asking how they know that the change in resistance is due only to the change in length.
> (Analysis of Classroom Passages – The Resistance of a Wire)

> K: They are evaluating not only the length, but also the type of wire. The type of wire will make the greater impact on the resistance and therefore any evaluations of the length's effects are clouded by the different types of wires.
> (Analysis of Classroom Passages – The Resistance of a Wire)

Nina:

> N: I would ask the students to use one type of wire in their experiment and not vary the material. The type of wire should be consistent.
> (Analysis of Classroom Passages – The Resistance of a Wire)

Conceptions Regarding Appropriate Manipulation of the Independent Variable

In the *Think-Aloud Experimental Design Interview* (inclined plane context), the participants were asked to compose an outline of an experiment to investigate the relationship between the minimum force necessary to pull a block up an inclined plane and the angle of the inclined plane surface. The experiment was to be conducted using only the equipment provided.

In their experimental design outline, each participant requested equal intervals between independent variable values but did not request that applied force measurements be taken at angles above 50 degrees. The following interview excerpts include the participants' descriptions of how they planned to manipulate the independent variable (angle of inclination):

<u>Betty</u>

> B: I think I would prefer using intervals of 5 degrees or so. So, I would have the students do a trial at 5, 10, 15,… all the way up to say 40 degrees.
> *(Think-Aloud Experimental Design Interview – The Inclined Plane)*

<u>Kurt</u>:

> K: I'd have them take a measurement every 10 degrees starting at zero and ending at around 50 degrees. That should be enough to get a good line.
> *(Think-Aloud Experimental Design Interview – The Inclined Plane)*

<u>Nina</u>:

> N: I think if they go up to 45 degrees they'll see the relationship. Maybe a measurement every fifteen degrees along the way. Then they can just extend the line.
> *(Think-Aloud Experimental Design Interview – The Inclined Plane)*

After the participants described their experimental designs both verbally and in writing, they were asked to indicate their expectations regarding the relationship between angle of inclination and minimum applied force. All three participants indicated that they expected a linear relationship.

In response to the hypothetical scenario (item 4 in Figure 5.2), the participants again (collectively) focused on the need for both equal intervals between and an adequate range of independent variable (angle of inclination) values. The following excerpts contain the participants' responses to this item.

<u>Betty</u>:

> B: It would probably be more beneficial to them to have a more steady interval. I would ask them why they jumped from 5 degrees to 65 degrees. So it would be ok for them to start at five but then perhaps pick some smaller interval to get to their next reading and continue using that interval for the next few trials.
> *(Analysis of Classroom Passages – The Inclined Plane)*

Kurt:

> K: I would expect group B (inclined plane context) to fully illustrate the relationship because of the consistent interval and large range of angles. Group A does not keep a consistent interval for their data and therefore until we graph the data, it is much more difficult to see the relationship.
> *(Analysis of Classroom Passages – The Inclined Plane)*

Nina:

> N: I would ask Group A and Group C why they picked those values. Group A's range is too broad (5 degrees – 65 degrees) and Group C their range is too close. The angles should be more different.
> I: And so you don't like group C's range of one to five degrees.
> N: No, I think we need a little greater...I think just a little bit of greater range to see the pattern.
> *(Analysis of Classroom Passages – The Inclined Plane)*

When asked to provide a rationale for their preference for equal intervals between and an adequate range of independent variable values, the participants described issues related to both students' graph interpretation and to instrument precision. The following excerpts contain portions of the participants' rationales.

<u>Betty</u>:

B: Umm...because if you're just going from one to five or five to ten (degrees), it's just a very little tiny part of the graph…you might miss a lot that's going on around it.
(Analysis of Classroom Passages – The Inclined Plane)

I: So what if it (the relationship) was supposed to be non-linear and there was that large interval between measurements?
B: Then you wouldn't see all the stuff in between…like places where it (the graph) curves.
(Analysis of Classroom Passages – The Inclined Plane)

<u>Kurt</u>:

K: When graphing our data, if we have a large gap in our data, we are unsure of how the data is represented in the interval and our estimation from the data at the ends of this gap may or may not give us an accurate account of what's going on in the gap.
(Analysis of Classroom Passages – The Inclined Plane)

<u>Nina</u>:

N: If there isn't a whole lot of distinction between them (a narrow range of independent variable values), the spring scale may not be able to measure a difference in force.
(Analysis of Classroom Passages – The Inclined Plane)

N: I think the values are too close. I think that they're not gonna get enough accuracy with the spring scales. You wouldn't be able to see a whole lot of, you know, difference in the force with that small. I think, you know, varying in tens (of degrees) or so…maybe that would be a lot better.
(Analysis of Classroom Passages – The Inclined Plane)

Discussion

The participants’ responses to prompts related to controlled experimentation suggest that each recognizes the purposes of controlling variables. However, the data also suggest that one’s ability to recognize uncontrolled experimentation does not assure that he or she will have the ability to *design* a controlled experiment. For example, in the electricity context, Nina recognized an uncontrolled experiment involving the concurrent modification of both wire resistivity (material) and length. In earlier responses, Nina did not identify either length or material as being influential to the resistance of a wire nor did she plan to control these variables in her own experimental design. It appears that Nina’s conceptions about resistivity were not well formed, not allowing her to see the importance of length and material. A similar relationship between controlling variables and one’s subject-specific science concepts was observed in other studies as well (e.g., Lawson, 1985; Levine & Linn, 1977; Linn & Swiney, 1981; Schauble, Klopfer, & Raghavan, 1991).

When responding to the inclined plane-based scenario, the participants consistently expressed a preference that the independent variable be manipulated such that its values are separated by a consistent interval. For example, Kurt indicated that a consistent interval between measurement values would help students interpret the resulting graph. In addition, the participants recommended that the independent variable be manipulated such that its relationship with the dependent variable would be thoroughly illustrated. For instance, Nina indicated that a sufficient range of data should be collected to ensure that the students will “see the relationship.” In general, the participants’ rationales were based mainly upon a projected need to easily construct and interpret graphical representations.

Interestingly, the actual experiment designed by the participants did not require that the independent variable be manipulated in such a way that an adequate range of dependent variable values be obtained. That is, none of the participants planned to measure the applied force when the surface of the incline was tilted above an angle of 50 degrees. This tendency seemed to make sense since each participant suspected that students could recognize a linear relationship from a small number of ordered pairs and could extrapolate from the resulting trend line. But the relationship is not actually linear and as a result of the small range of measurements required in the participants' experimental designs, it is not likely that the participants' students would have recognized the non-linearity of the relationship between minimum applied force and angle of inclination (depending on the coefficient of kinetic friction, the non-linear nature of the relationship starts to become evident at angles of approximately 65 degrees and above).

In sum, the data from the hypothetical classroom scenarios indicated that each participant recognized uncontrolled experimentation and inappropriate manipulation of the independent variable. However, in designing their own experiments, none of the participants controlled all pertinent variables or planned to manipulate the independent variable in a way conducive to fully illustrating the nature of its relationship with the dependent variable. Clearly, the participants' subject-specific physics concepts (e.g., those related to the resistivity of a wire or inclined plane mechanics) influenced their responses to the task-based protocols to a greater degree than they did the hypothetical scenarios. This suggests that the task-based protocols used in this study tended to assess more complex and integrated understandings of experimentation than did the hypothetical scenarios.

The findings of this study also suggest the possibility that completely different types of understandings are used to conduct experiments and react to hypothetical scenarios. Additional evidence of this difference can be found in other science education studies. For example, Brumby (1982) studied medical students' understanding of the nature of living things by giving them a rock and asking them how they could determine whether the rock had ever been alive. Brumby indicated that many of the medical students had difficulty planning appropriate scientific experiments and went on to suggest that these students would have likely performed better in response to pictures or drawings.

White and Gunstone (1992) attributed the observed differences between students' thinking about objects and their thinking about pictorial representations in part to differing levels of abstraction among research protocols.

They suggested that:

> Drawings are an abstraction, and may signal to the student that what is wanted is a response from abstract knowledge, where objects call for a response from concrete experience. Drawings and objects may tap different forms of understanding, and reveal inconsistent beliefs that a student holds (p. 75).

One possible explanation for the inconsistent conceptions of experimentation demonstrated in this study is that the participants' reactions to data described in hypothetical scenarios tapped a different form of understanding than those elicited by designing an experiment with actual equipment and thinking out-loud about the process.

Implications for Future Research

The discrepancies observed in the participants' responses to *seemingly* similar research prompts highlights the utility of using multiple types of research protocols. In this study, using both hypothetical scenarios *and* think-aloud interviews proved to be useful in that it provided insights that might not have been visible using either of the protocols in isolation. This combination of protocols suggests that similar research might also utilize multiple research methods and/or protocols. Schoenfeld (1985) summarized the need for using multiple research protocols:

> Any experimenter who wishes to get a sense of the whole cognitive picture should consider using a range of complementary (verbal and other) methodologies, and must be extremely cautious in interpreting the results obtained from a body of methodologically similar studies (p. 174).

Therefore, future investigators may wish to expand even further the protocols used in this study by designing additional ways to elicit their participants' conceptions of experimentation.

One such protocol might capitalize on the tendency of the participants' in this study to include questions in their reactions to the student-generated procedures or data presented in the hypothetical scenarios. White and Gunstone (1992) described a powerful research protocol called *question production* that, in one variation, requires research participants to generate questions in response to a *stimulus*. White and Gunstone identified graphs or tables of data as possible stimuli.

The hypothetical scenarios used in this study incorporated similar stimuli but did not explicitly request that the participants formulate questions to pose to the students. As mentioned previously, the participants in this study sometimes posed their reactions as questions without any prompting from the authors. At times, examining these questions helped clarify or reinforce the interpretation of the participants’ understandings (e.g., the question Kurt planned to pose regarding uncontrolled experimentation). The use of protocols that explicitly require the formulation of such questions would have likely enriched the data collected in this study and should be explored in future research.

White and Gunstone (1992) suggested that teachers should attempt to ask questions that will probe their students’ understandings. Such questions, sometimes referred to as “thinking questions,” typically require students to apply, extend, or reinterpret existing ideas. White and Gunstone also concluded that the ability to formulate thinking questions is an important indication of the depth of a teacher’s knowledge base. Specifically, formulating questions that are effective in probing understanding requires teachers to possess deep understandings of student learning and of the subject matter.

That said, the use of *question production* protocols to examine science teachers’ ability or tendency to ask thinking questions is likely to make a substantial contribution to what science teacher educators have learned about teacher knowledge and development. Therefore, teacher education researchers should continue to investigate the usefulness of task-based and scenario-based protocols as well as begin to emphasize the use of question production-type protocols in their efforts to diversify the research protocols currently in use.

CHAPTER 6

SECONDARY SCHOOL PHYSICS TEACHERS' CONCEPTIONS OF SCIENTIFIC EVIDENCE: SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

Summary

The primary purpose of this study was to describe and interpret the nature of prospective and practicing physics teachers’ conceptions of the measurement reliability and experimental validity of scientific evidence. Specifically this research aimed to address the following research questions:

- What types of issues related to the measurement reliability and experimental validity of scientific evidence do practicing and prospective physics teachers think about when designing experiments?
- When presented with hypothetical scenarios that describe unsound experimental procedures or poorly supported conclusions (or both), what concerns will prospective and practicing physics teachers raise?
- When the participants’ responses to parallel research prompts are compared across protocols, what similarities and differences exist?

This section will summarize briefly what was learned concerning each of these three questions.

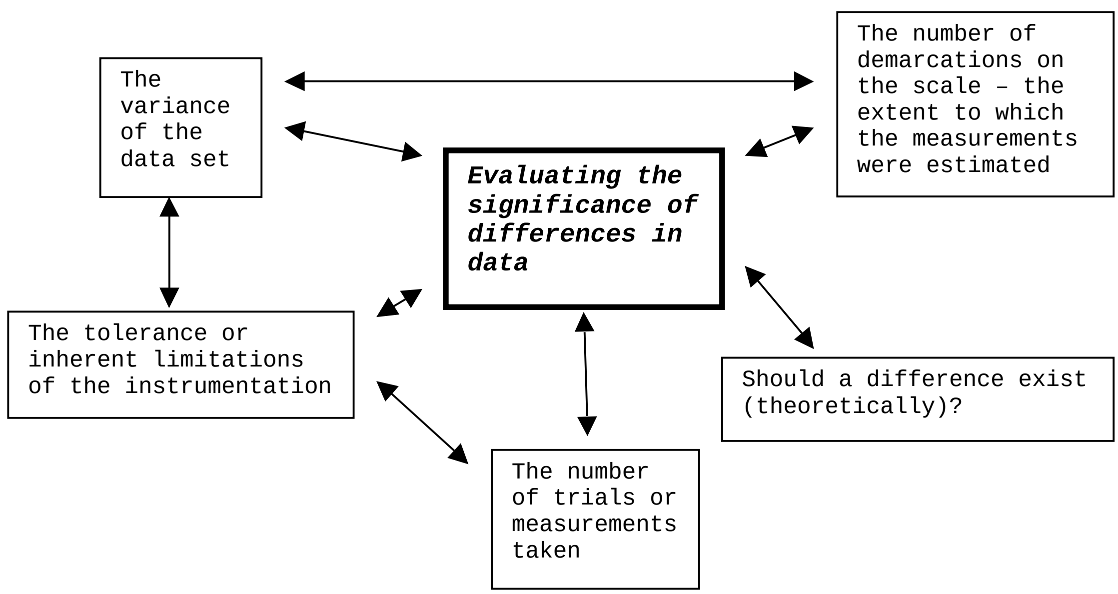
An analysis of the data collected in this study suggests that each participant's ability to critically evaluate scientific evidence was influenced by his or her ability to integrate conceptions of scientific evidence and subject-specific science concepts. For example, the participants' evaluation of whether or not a spring balance truly measured the minimum force necessary to pull a block up an incline was supported by knowledge of both instrumentation and inclined plane mechanics. These types of relationships were identified in the responses of all three participants. Table 6.1 summarizes the nature of the relationship found between conceptions of scientific evidence and subject-specific science concepts.

Table 6.1. Conceptions of scientific evidence: Integration with subject-specific science concepts

More Integrated	**Less Integrated**
· Controlled Experimentation · Generalization of Conclusions · Measurement Validity · Statistical significances of differences in data · Recognition and Treatment of Outliers · Instrument Choice (scale and precision · Manipulation of independent variables	· Control of Variables · Rationale for Repeated Trials · Reliability of Data

The integration of ideas observed in this study was not limited solely to the integration of conceptions of scientific evidence with pertinent subject-specific science concepts. Especially in prompts that required an evaluation of the significance of differences in data, multiple conceptions of scientific evidence were integrated with each other. This complex use of ideas was particularly evident in Kurt's responses. Figure 6.1 is a graphical representation of the participants' collective criteria for evaluating the significance of differences in data.

Figure 6.1. The participants' collective ideas regarding the evaluation of differences in data



The study also revealed how the necessity to integrate appropriate subject-specific physics concepts varied across research protocol. Specifically, the participants often recognized faulty experimentation even in the absence of important subject-specific physics concepts (see Table 6.2). The participants' alternative subject-specific physics conceptions were much more influential in their responses to the task-based interview protocols.

Table 6.2. Conceptions of experimentation across research protocol

Target Conception	Nature of the participants' responses to the Analysis of Classroom Passages Protocol	Nature of the participants' responses to the Think-Aloud Experimental Design Interview Protocol
Controlled Experimentation	· Uncontrolled experiments recognized immediately · Comparisons impossible unless variables controlled	· Alternative subject-specific physics conceptions revealed · Uncontrolled experiments designed
Independent Variable Manipulation	· Exceptionally large or small intervals between dependent variable measurements recognized immediately · Inadequate ranges of measurements recognized immediately	· Independent variables manipulated such that an inadequate range of dependent variables would result · Directly linked to naïve expectation of linear relationship between variables

Implications for Future Research

An important issue related to this study concerns whether or not differences in the participants' responses could be connected directly to differences in their teaching experience. Establishing this connection was not a focus of this study but had it been, I would have needed to establish more similarities in the participants' backgrounds. Identifying such similarities would have required that I collect much more demographic information. Therefore, future research aimed at making connections between conceptions of scientific evidence and teaching experience might analyze demographic data similar to that collected in this study plus additional participant data such as: their engagement in professional communities (e.g., *The American Association of Physics Teachers*), the nature of the curriculum they use (e.g., inquiry-oriented or more traditional), and their goals for teaching physics.

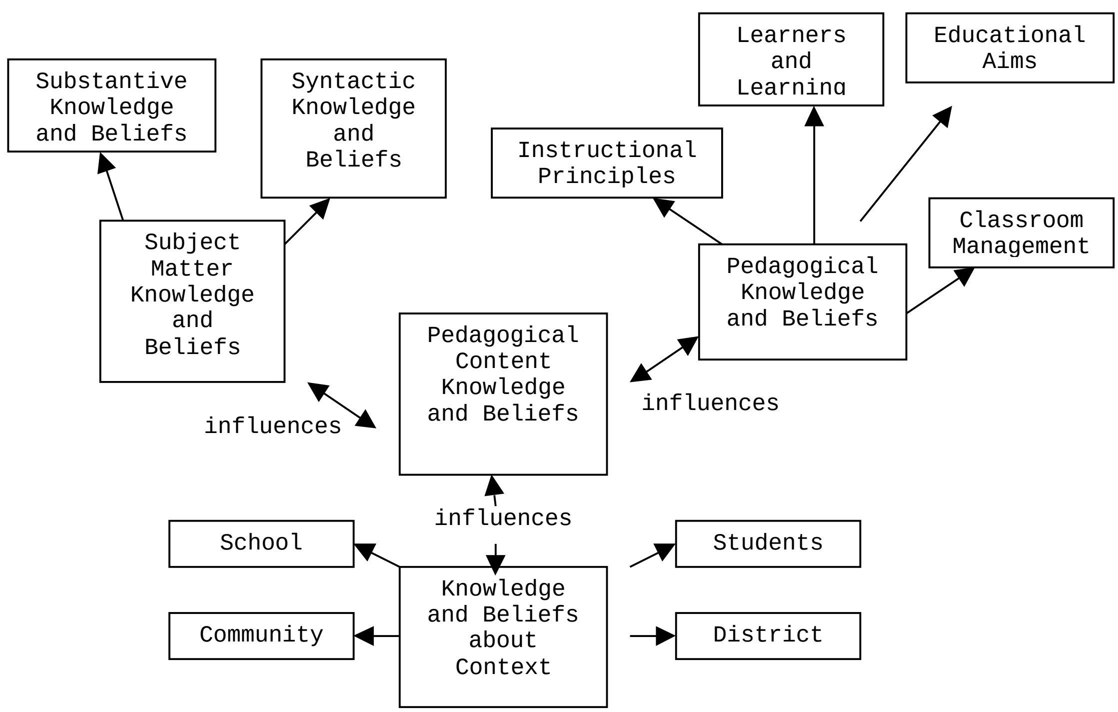
The diversification of future research data need not be limited to demographics. Future researchers' might design additional protocols that lend further insight into physics teachers' conceptions of scientific evidence. The design of future protocols, especially those that are scenario-based, should be informed by what transpired in this study. For example, future researchers should be aware that prompts intended to examine respondents' identification of experimental design flaws (e.g., flaws related to measurement validity or the generalization of findings) may yield limited data. In this study, the experimental design flaws present in many of the hypothetical scenarios were somewhat subtle and were not immediately identified by the participants. In this instance, the researcher must either be satisfied with the "everything seems fine" response or decide how to focus the participant on important issues related to the hypothetical experiment. I found that the latter strategy allowed me to probe deeper into the participants' understandings. Consequently, if future researchers are to use protocol that examine the identification of experimental design flaws, I recommend that these protocols be administered in a "one-on-one" interview format so that the researcher may ask for clarification, focus, and elaboration.

Implications for Physics Teacher Education

In this study, the participants often integrated subject-specific physics concepts with concepts of scientific evidence such as those related to instrumentation, experimental design, and statistical significance to evaluate the measurement reliability or experimental validity of scientific evidence. In Schwab's (1964, 1978) terminology, the participants frequently integrated aspects of both syntactic and substantive subject-matter knowledge while evaluating scientific evidence.

Numerous teacher education scholars have identified substantive and syntactic subject matter knowledge as influential to other domains of knowledge that are critical to effective teaching (e.g., Borko & Putnam, 1996, Grossman, 1990; Magnusson, Krajcik, & Borko, 1999; Shulman, 1986). One of the knowledge domains linked to substantive and syntactic subject matter knowledge is *pedagogical content knowledge.* Shulman (1986) suggested that pedagogical content knowledge represents an important knowledge domain for teaching and described it as knowledge of "the ways of representing and formulating the subject that makes it comprehensible for others" (p. 9). Similarly, Magnusson, Krajcik, and Borko (1999) viewed pedagogical content knowledge as a knowledge domain unique to teaching that results from the transformation of knowledge from several other domains (e.g., syntactic subject matter knowledge, pedagogical knowledge – see Figure 6.2).

Figure 6.2. A model of the relationships among the Domains of Teacher Knowledge



Source: Adapted from Magnusson, Krajcik, and Borko (1999)

This model suggests that teachers' pedagogical content knowledge is critically influenced by their understandings of the substantive and syntactic aspects of the discipline. Therefore, if physics teachers are to transform their subject matter knowledge of physics in a way comprehensible to students, they must draw upon a well-developed substantive and syntactic knowledge base.

This raises and important question for physics teacher educators: what types of instructional interventions are likely to help prospective and practicing physics teachers develop both substantive and syntactic knowledge? I believe the findings of this study provide insight into this question. Since the participants in this study often integrated substantive and syntactic knowledge to evaluate scientific evidence, it is clear that instructional interventions intended to help physics teachers develop substantive and syntactic knowledge should nurture concept integration and not discourage it. Science teacher educators might consider avoiding "general" approaches (Ennis, 1989) to helping physics teachers develop critical thinking abilities. According to Ennis (1989), a general approach "attempts to teach critical thinking abilities and dispositions separately from the presentation of the content" (p. 4). Many of these instructional efforts were deemed unsuccessful because the critical thinking abilities did not transfer across contexts (e.g., Pressley, Snyder, & Cariglia-Bull, 1987; Salomon & Globerson, 1987). Some researchers have attributed this problem to the assertion that critical thinking is too subject-matter knowledge sensitive to be taught in this way (Norris, 1985), while others attribute the problem to instruction that was divorced from meaningful context (Schauble, Glaser, Duschl, Schulze, & John 1994).

Instead, the effective partnership between substantive and syntactic knowledge that was demonstrated in this study could become a focus in pre-service and in-service physics teacher education. This suggestion incorporates the recommendations of numerous scholars and researchers who have recommended a renewed emphasis on substantive knowledge in science teacher education (e.g., Abd-El-Khalick & Boulaoude, 1997; Lederman & Latz, 1995) and extends them to a vision of physics teacher education where the interplay of substantive and syntactic knowledge is encouraged and nurtured.

The findings of this study also indicate that the participants' involvement of authentic, subject-matter (both substantive and syntactic) knowledge-intensive tasks such as those required in the *Think-Aloud Experimental Design Interviews* often encouraged a more complex line of reasoning about scientific evidence. This strongly suggests we include research-related tasks in physics teacher education. The following section describes more fully how physics teachers' participation in such tasks might become more visible in both pre-service and in-service teacher education.

Recommendations for Program Design in Physics Teacher Education

Original Research

Some science teacher educators have described what curriculum and instruction aimed at the development of both substantive and syntactic subject-matter knowledge might look like. Friedler and Tamir (1986) developed an instructional module for use with high school biology students called *Basic Concepts of Scientific Research*. In this module, students discussed selected issues that relate to knowledge construction in science. The module culminated with the students conducting original (at least to them) scientific research, which helped the students integrate science content knowledge with the key epistemological concepts of the module. Tamir (1989) described the positive results of the module and suggested that similar instruction could be designed for prospective science teachers. Based upon their study of undergraduate physics students’ conceptions of measurement errors and statistics, Sere, Journeaux, and Larcher (1993) also supported an integrated approach suggesting that instructional interventions concurrently address issues of subject-specific physics *and* data evaluation.

The involvement of prospective and practicing science teachers in original scientific research has been recommended by a growing number of scholars and researchers (e.g., Grossman, Wilson, & Shulman, 1989; Haefner, 2001; Loucks-Horsley, et al., 1998; van Tilburg, Verloop, & Vermunt, 1999). Specifically, Grossman, Wilson, and Shulman (1989) suggested that: “In learning to conduct their own inquiries-scientific, historical, mathematical, literary, or otherwise-students learn the difference between evidence that is acceptable and unacceptable, sufficient and insufficient” (p. 30). Similarly, Gess-Newsome (1999) described original research as an activity that can help teachers become familiar with the nature of knowledge construction and validation in a respective field.

The practice of encouraging students to conduct original research is visible in colleges of science sometimes does not occur until graduate school or at the advanced stages of a baccalaureate program (McDermott, 1990a; National Science Foundation, 1996). In some physics teacher education programs, advanced level coursework centered on original research is not required since only a limited number of credit hours can be devoted to physics content preparation (Tobias, 1992) and limited human and physical resources support such research. Consequently, some prospective physics teachers complete their undergraduate programs without ever engaging in original scientific research. This reality emphasizes the importance of incorporating original research into both physics teaching methods courses and in-service teacher professional development experiences.

Such courses or professional development experiences might take the form of undergraduate or graduate colloquia designed primarily as an opportunity for physics teachers to collaborate; they can immerse themselves in a small number of original investigations related to their respective field(s) of interest. These investigations should span several content areas within a given discipline to increase the likelihood that any concept integration would take place in unique or diverse ways. This vision of prospective and practicing teacher collaboration is consistent with established strategies for science teacher professional development (see *Study Groups* – Loucks-Horsley et al., 1998).

The development of undergraduate/graduate research colloquia might provide a forum for meaningful collaboration between faculty in colleges of science (or scientists from non-academic settings) and colleges of education. For scientists, working with prospective and practicing physics teachers might be seen as a worthwhile aspect of their community service. Further, there is mounting evidence that science teachers' research experiences are enriched under the mentorship of a scientist in their field. For example, Gilmer (1999) studied the K-8 science teachers who participated in the *Teachers Learning Inquiry through Scientific Research Project* at Florida State University. This project engaged teachers in original scientific research under the guidance of scientists from both academic and non-academic settings. Gilmer's study of the teachers who participated in this project revealed that many "had an opportunity to practice methods of science, experience inquiry, use the language of science, both in writing and in speaking, and they came to understand the culture of science” (p. 22). An equally important finding of this study is that most of the teachers returned to their classrooms after the project and provided similar learning opportunities to their students, something they had felt uncomfortable doing in the past. Similarly, prospective and practicing physics teachers are more likely to engage their students in original research if they have become comfortable with conducting research themselves.

Evidence-Based Argumentation and the Discourse of Peer Review

It is unlikely, however, that the act of conducting original research, in and of itself, will fully support the development of substantive and syntactic understandings (Haefner, 2001). Therefore, teacher education courses or the aforementioned undergraduate/graduate colloquia might encourage physics teachers to augment their research experiences with activities that simulate other aspects of the scholarly work of scientists. One such activity involves teachers using the data collected during their research to construct an evidence-based argument that is presented to peers for review and critique.

This recommendation was informed by the findings of this study as well as the recommendations of other scholars and researchers (e.g., Boulter & Gilbert, 1995; McDermott, 1990a; Pea & Greeno, 1990). For example, McDermott (1990a) suggested that physics teachers learn not only physics concepts but also the evidence and reasoning that were used in developing those concepts. Lampert (1990) noted that classroom culture should encourage discourse among peers regarding the deliberations, problems, risks, and issues that underlie the production of scientific knowledge. Latour and Woolgar (1986) connected these goals specifically to argument construction when they suggested that the process of constructing an argument encourages one to weigh evidence/data, assess alternative explanations, and evaluate the viability of scientific claims. It is likely that one hones these critical thinking abilities while constructing and preparing an argument for peer review, as well as reviewing the arguments of others.

Evidence of meaningful student learning as a result of argument/peer review-based instruction has been documented in research with children (Bell, 1998; Brown & Campione,

1990) and high school students (Sandoval & Reiser, 1997; Tabak & Reiser, 1999). Brown and Campione (1990) concluded that the discourse that evolved in their research setting helped promote “significant improvements in the students’ thinking skills and in the domain-specific knowledge about which they are reasoning” (p. 124). Although the use of argument/peer review-based instruction with prospective and practicing teachers has received little attention in the science teacher education literature, the *lack* of its use with teachers has been recognized (e.g., Smith, Conway, & Levine-Rose, 1995). The apparent disregard for argumentation in physics teacher education is of great concern since the failure to engage K-12 students’ in argumentation has been associated with the inability to critically evaluate scientific claims (Norris & Phillips, 1994; Solomon, 1991).

Sparks (1997) suggested that it is very difficult for science teachers to support a type of learning that they have never experienced themselves. Therefore, to better support students in argumentation and discourse about scientific evidence, prospective and practicing science teachers should also engage, from the perspective of the science learner, in argument construction and presentation.

An instructional environment that emphasizes both the construction and critical evaluation of evidence-based arguments will likely encourage a more accurate vision of scientific knowledge construction (one where discursive practices are seen as integral to the process) than that portrayed by confirmatory laboratory courses so visible in many undergraduate physics programs (Driver, Newton, & Osborne, 2000; Lampert, 1990). For example, amidst the diversity of data and interpretations that may emerge from the investigations, there will most likely be contradictory interpretations presented. If the “right answers” to the research questions or hypotheses are not known or emphasized, educators can engage practicing and prospective physics teachers in post-presentation negotiations aimed at resolving conflicts. Many in the science education community have identified this type of argumentation and negotiation as similar to the cooperative construction of knowledge that often takes place in the expert scientific community (e.g., Carey, 1985; Kuhn, 1962; Nersessian, 1989). Further, post-presentation negotiations are likely to help illustrate the theory-laden nature of observation, common logical fallacies in argumentation, and the characteristics of compelling evidence.

In addition to providing the science teacher educator with valuable information about prospective and practicing physics teachers' conceptions of the scientific evidence and other epistemological conceptions, the nature and quality of the arguments presented will likely indicate understandings of subject-specific physics concepts. When the learner engages in argumentation, he or she divulges the *status* of the science concepts that are used in the argument (Posner, Strike, Hewson, & Gertzog, 1982). That is, the degree to which the learner views the concept as intelligible, plausible, or fruitful is made more explicit. Thus, discourse centered upon scientific phenomenon can be thought of as the *negotiation of concept status.* Many conceptual change scholars have cited the importance of concept status negotiations in promoting conceptual change (e.g., Hewson, Beeth, & Thorley, 1998; Smith, 1987). Therefore, the process of argument presentation and critique might also be useful in informing science teacher educators in their design of further instructional interventions aimed at the development of appropriate subject-specific science concepts as well as appropriate conceptions of scientific evidence.

Multiple Opportunities to Evaluate Scientific Evidence

Since the ability to critically evaluate evidence has been linked to practice (Driver, Newton, & Osborne, 2000; Greeno, 1992; Sandoval & Reiser, 1997; Tabak & Reiser, 1999), science teacher educators might also provide practicing and prospective physics teachers with additional opportunities to evaluate scientific evidence by utilizing classroom passages similar to those used in this study. The evidence presented in these hypothetical scenarios could be the subject of both individual and large group analyses.

It should be noted, however, that these additional opportunities do not have to be limited to the evaluation of hypothetical student-generated evidence; science teacher educators could also provide practicing and prospective physics teachers with the opportunity to evaluate scientific evidence collected by actual secondary school students. To do this, science teacher educators could call upon the resources and expertise of secondary school physics teachers from surrounding communities. Science teacher educators could work along side these teachers in an effort to convert evidence and data from “real” students’ investigations into a format appropriate for inclusion in teacher education curriculum.

Collaboration surrounding the evaluation of students’ laboratory work has been strongly recommended in the professional development literature (e.g., Feldman, 1993; Loucks-Horsley et al., 1998; NRC, 2000). Feldman (1993) studied the physics teachers involved in the *Physics Teacher Action Research Group* (PTARG), a group of San Francisco area physics teachers who met regularly to discuss and collaboratively evaluate their students’ work. Feldman found that the teachers in this group, because of their collaboration, developed both pedagogical content knowledge as well as a deeper understanding of subject-specific physics concepts.

Adaptation of Evidence-Based Arguments for Publication in Refereed Journals

Another approach that science teacher educators can take to simulate the scholarly work of scientists is encouraging teachers to adapt aspects of their arguments (e.g., research methods, findings) for publication in an appropriate refereed journal. This activity would encourage physics teachers to develop content understandings that would allow them to situate their research in the findings of others. In addition, preparation of a scholarly manuscript would help physics teachers become familiar with established norms for describing procedures and instrumentation, reporting measurements, and choosing and representing statistical tests.

Conclusion

This chapter has described a multi-faceted vision of physics teacher education that includes the infusion of physics teaching methods courses or undergraduate/graduate research colloquia with experiences intended to simulate the scholarship of scientists. Specifically, this vision recommends experiences such as conducting original research, constructing evidence-based arguments, critically reviewing others' arguments, evaluating students' laboratory work (similar to the PTARG model), and generating publishable research reports. These research experiences are recommended in light of the existing research literature, which links these types of experiences to the potential for developing science teachers' substantive and syntactic subject-matter knowledge.

These types of subject-matter knowledge are clearly desirable if physics teachers are to implement reform-based teaching. More specifically, science teachers' substantive and syntactic understandings have been associated with their ability to support students in discourse surrounding scientific evidence. However, this emphasis on the development of subject-matter knowledge extends beyond the need to implement reform-based physics instruction. It is likely that the physics teacher who develops these knowledge bases will be better able to critically evaluate the experimental data or conclusions reported in his or her scientific field and consequently, will be more likely to stay abreast of new developments in the field. Therefore, physics teacher education programs should include instruction aimed at the development of both substantive and syntactic knowledge as a means of obtaining their global objective of supporting physics teachers as lifelong learners.

References

Abd-El-Khalick, F., & Boulaoude, S. (1997). An exploratory study of the knowledge base for science teaching. Journal of Research in Science Teaching, 34, 673-699.

Abell, S., & Smith, D. (1994). What is science? Preservice elementary teachers' conceptions of the nature of science. International Journal of Science Education, 16, 475-487.

Aguirre, J. M., Haggerty, S. M., & Linder, C. J. (1990). Student teachers' conceptions of science, teaching and learning: A case study in preservice science education. International Journal of Science Education, 12, 381-390.

Aikenhead, G. (1987). High school graduates beliefs about science-technology-society: Characteristics and limitations of scientific knowledge. Science Education, 71, 459-487.

Aikenhead, G. S., & Ryan, A. (1989). The development of a multiple choice instrument for monitoring views on science-technology-society topics (Research Report), Social Sciences and Humanities Research Council of Canada, Ottowa, Canada.

Allen, R. D., Statkiewitz, W. R., & Donovan, M. (1983). Student perceptions of evidence and interpretations. In J. Novak (Ed.), Proceedings of the international seminar: Misconceptions in science and mathematics (pp. 79-83). Ithaca, NY: Cornell University.

Allie, S., Buffler, A., Kaunda, L., Campbell, B., & Lubben, F. (1998). First-year physics students' perceptions of the quality of experimental measurements. International Journal of Science Education, 20, 447-459.

American Association for the Advancement of Science. (1989). Science for All Americans. New York: Oxford University Press.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. New York: Oxford University Press.

American Association for the Advancement of Science. (2001). Atlas of Science Literacy. Washington, DC: American Association for the Advancement of Science/National Science Teachers Association.

Ball, D. L. (1988). Unlearning to teach mathematics (Issue Paper 88-1). East Lansing: Michigan State University, National Center for Research on Teacher Education.

Bauer, H. H. (1992). Scientific literacy and the myth of the scientific method. Chicago: University of Illinois Press.

Baxter, G. P., Shavelson, R. J., Goldman, S. R., & Pine, J. (1992). Evaluation of procedure-based scoring for hands-on science assessment. Journal of Educational Measurement, 29 (1), 1-17.

Bell, P. L. (1998). Designing for students' science learning using argumentation and classroom debate. Unpublished doctoral dissertation, University of California, Berkeley.

Bogdan, R. C., & Biklen, S. K. (1998). Qualitative research for education: An introduction to theory and methods. Boston: Allyn and Bacon.

Borko, H., & Putnam, R. T. (1996). Learning to teach. In D.C. Berliner & R. C. Calfee (Eds.), Handbook of educational psychology (pp. 673-708). New York: Macmillan.

Boulter, C. J., & Gilbert, J. K. (1995). Argument and science education. In P. Costello & S. Mitchell (Eds.), Competing and consensual voices: The theory and practice of argument (pp. 84-98). Philadelphia: Multilingual Matters LTD.

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking, or a context by any other name. In D. Kuhn (Ed.), Developmental perspectives on teaching and learning thinking skills (pp. 108-125). Basel (Switzerland): Karger.

Brumby, M. (1982). Medical students' perceptions of science. Research in Science Education, 12, 107-114.

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: MIT Press.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try it and see if it works": A study of grade 7 students' understanding of the construction of scientific knowledge. International Journal of Science Education, 11, 514-529.

Carlsen, W. (1991). Subject-matter knowledge and science teaching: A pragmatic perspective. In J. E. Brophy (Ed.), Advances in research on teaching (Vol. II. Teachers' subject-matter knowledge and classroom instruction, pp. 115-186). Greenwich, CT: JAI.

Case, R. (1974). Structures and strictures: Some functional limitations on the course of cognitive growth. Cognitive Psychology, 6, 544-573.

Champagne, A. B., Klopfer, L. E., & Gunstone, R. F. (1982). Cognitive research and the design of science instruction. Educational Psychologist, 17, 31-53.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research, 63, 1-49.

Clement, J., (1982). Students' preconceptions in elementary mechanics. American Journal of Physics, 50, 66-71.

Creswell, J. W. (1998). Qualitative inquiry and research design: Choosing among five traditions. London: Sage Publications.

DeBoer, G. E. (1991). A history of ideas in science education: Implications for practice. New York: Teachers College Press.

Desautels, J. & Larochelle, M. (1998). The epistemology of students: The 'thingified' nature of scientific knowledge. In B. J. Fraser & K. G. Tobin (Eds.), International Handbook of Science Education (pp. 115-127). Dordrecht: Kluwer Academic Publishers.

Dewey, J. (1933). How we think. New York: D.C. Heath.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young people’s images of science. Philadelphia: Open University Press.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. Science Education, 84, 287-312.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). Making sense of secondary science: Research into children’s ideas. London: Routledge.

Duggan, S., Johnson, P., & Gott, R. (1996). A critical point in investigative work: Defining variables. Journal of Research in Science Teaching, 33, 461-474.

Duschl, R. A. (1990). Restructuring science education: The importance of theories and their development. New York: Teachers College Press.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. Educational Researcher, 18 (3), 4-10.

Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). Doing naturalistic inquiry: A guide to methods. London: Sage.

Feldman, A. (1993). Teachers learning from teachers: Knowledge and understanding in collaborative action research. Unpublished doctoral dissertation. Stanford University.

Finegold, M., & Gorsky, P. (1991). Students' concepts of force as applied to related physical systems: A search for consistency. International Journal of Science Education, 13, 97-113.

Foulds, K., Gott, R., & Feasey, R. (1992). Investigative work in science. Unpublished research report, University of Durham.

Friedler, Y., & Tamir, P. (1986). Teaching basic concepts of scientific research to high school students. Journal of Biological Education, 20, 263-270.

Garfield, J. & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. Journal of Research in Mathematics Education, 19, 44-63.

Gess-Newsome, J. (1999). Secondary teachers' knowledge and beliefs about subject matter and their impact on instruction. In J. Gess-Newsome & N. G. Lederman (Eds.), Examining pedagogical content knowledge: The construct and its implications for science education (pp. 51-94). Dordrecht: Kluwer.

Gilmer, P. J. (1999). Teachers learning inquiry through scientific research. In T. L. Keilborn & P. J. Gilmer (Eds.), Meaningful science: Teachers doing inquiry and teaching science (pp. 11-26). Tallahassee, FL: SERVE.

Glaser, B., & Strauss, A. (1967). The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine.

Gott, R., & Duggan, S. (1995). Investigative work in the science curriculum. Buckingham: Open University Press.

Gott, R., & Duggan, S. (1996). Practical work: its role in the understanding of evidence in science. International Journal of Science Education, 18, 791-806.

Greeno, J. G. (1992). Mathematical and scientific thinking in classrooms and other situations. In D. F. Halpern (Ed.), Enhancing Thinking Skills in the Sciences and Mathematics (pp. 39-61). Hillsdale, NJ: Erlbaum.

Gronlund, N. E. (1998). Assessment of student achievement. Boston: Allyn and Bacon.

Grossman, P. (1990). The making of a teacher: Teacher knowledge and teacher education. New York: Teachers College Press.

Grossman, P. L., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. C. Reynolds (Ed.), Knowledge base for the beginning teacher (pp. 23-36). New York: Pergamon.

Guba, E. G., & Lincoln, Y. S. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

Haefner, L. B. (2001). Engaging in science inquiry: Prospective elementary teachers' learning in an innovative life science course. Unpublished doctoral dissertation, The Pennsylvania State University, University Park.

Herron, M. D. (1971). The nature of scientific enquiry. School Review, 79, 171-212.

Hewson, P. W., Beeth, M. E., & Thorley, N. R. (1998). Teaching for conceptual change. In B. J. Fraser & K. G. Tobin (Eds.), International Handbook of Science Education (pp. 199-218). Dordrecht: Kluwer Academic Publishers.

Holsti, O. R. (1969). Content analysis for the social sciences and humanities. Reading, MA: Addison-Wesley.

Holton, G. (1988). Thematic origins of scientific thought: Kepler to Einstein. Cambridge, MA: Harvard University Press.

Huber, R. A., & Moore, C. J. (2001). A model for extending hands-on science to be inquiry-based. School Science and Mathematics, 100, 32-41.

Inhelder, B. & Piaget, J. (1958). The growth of logical thinking. New York: Basic Books.

Joshua, S. & Dupin, J. J. (1987). Taking into account student conceptions in instructional strategy: An example in physics. Cognition and Instruction, 4, 117-135.

Jungwirth, E. (1985). Science teaching and pupil avoidance of logical fallacies. South African Journal of Education, 5, 55-60.

Jungwirth, E. (1987). Avoidance of logical fallacies – a neglected aspect of science-education and science-teacher education. Research in Science and Technological Education, 5, 43-58.

Jungwirth, E. (1990). Science teachers' spontaneous, latent, or non-attendance to the validity of conclusions in reported situations. Research in Science and Technological Education, 8, 103-115.

Jungwirth, E., & Dreyfus, A. (1975). Concepts of enquiry: Magic words or intellectual skills? Journal of Research in Science Teaching, 12, 305-314.

Jungwirth, E., & and Dreyfus, A. (1990). Identification and acceptance of a posteriori causal assertions invalidated by faulty enquiry methodology: An international study of curricular expectations and reality. In D. Herget (Ed.), More history and philosophy of science in science teaching (pp. 202-211). Tallahassee, FL: Florida State University.

Jungwirth, E., & Dreyfus, A. (1992). After this, therefore because of this: One way of jumping to conclusions. Journal of Biological Education, 26, 139-142.

Kitchener, K. (1983, Fall). Educational goals and reflective thinking. The Educational Forum, 75-95.

Kitchener, K. S., & King, P. M. (1981). Reflective Judgement: Concepts of justification and their relationship to age and education. Journal of Applied Developmental Psychology, 2, 89-116.

Kubiszyn, T., & Borich, G. (1990). Educational Testing and Measurement. Glenview, IL: Scott, Foresman/Little, Brown.

Kuhn, D. (1989). Children and adults as intuitive scientists. Psychological Review, 96, 674-689.

Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kyle, W. C. (1980). The distinction between inquiry and scientific inquiry and why high school students should be cognizant of the distinction. Journal of Research in Science Teaching, 17, 123-130.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. American Educational Research Journal, 27 (1), 29-63.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific facts. Princeton, NJ: Princeton University Press.

Lawson, A. E. (1985). A review of research on formal reasoning and science teaching. Journal of Research in Science Teaching, 22, 569-617.

Lawson, A. E. (1995). Science teaching and the development of thinking. Belmont, CA: Wadsworth.

Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., & Kwon, Y. J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? Journal of Research in Science Teaching, 37, 81-101.

Lederman, N. G., & Latz, M. S. (1995). Knowledge structures in the pre-service science teacher: Sources, development, interactions, and relationships to teaching. Journal of Science Teacher Education, 6, 1-19.

Lederman, N. & O'Malley, M. (1990). Students' perception of tentativeness in science: Development, use, and sources of change. Science Education, 74, 225-239.

Lederman, N. G., Wade, P. D., Bell, R. L. (1998). Assessing the nature of science: What is the nature of our assessments? Science and Education, 7, 595-615.

Lemke, J. (1990). Talking science: Language, learning and values. Norwood, NJ: Ablex.

Levine, D. I., & Linn, M. C. (1977). Scientific reasoning ability in adolescence: Theoretical viewpoints and educational implications. Journal of Research in Science Teaching, 14, 371-384.

Lewis, E. L. & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. Journal of Research in Science Teaching, 31, 657-678.

Liem, T. L. (1987). Invitation to science inquiry, (2nd ed.). Chino Hills: Science Inquiry Enterprises.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Newbury Park, CA: Sage.

Linn, M. C., & Swiney, J. F. (1981). Individual differences in formal thought: Role of expectations and aptitudes. Journal of Educational Psychology, 73, 274-286.

Linn, M. C., Clement, C., & Pulos, S. (1983). Is it formal if its not physics? Journal of Research in Science Teaching, 20, 755-770.

Loucks-Horsley, S., Hewson, P. W., Love, N., Stiles, K. E., Dyasi, H. M., Friel, S. N., Mumme, J., Sneider, C. I., & Worth, K. L. (1998). Designing professional development for teachers of science and mathematics. Thousand Oaks, CA: Corwin Press.

Lubben, F., & Millar, R. (1994). A survey of the understanding of children aged 11-16 of key ideas about evidence in science. PACKS Research Paper 3. York: University of York.

Lubben, F., & Millar, R. (1996). Children's ideas about the reliability of experimental data. International Journal of Science Education, 18, 955-968.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), Examining pedagogical content knowledge: The construct and its implications for science education (pp. 95-132). Dordrecht: Kluwer.

McDermott, L. C. (1990a). A perspective on teacher preparation in physics and other sciences: The need for special science courses for teachers. American Journal of Physics, 58, 734-742.

McDermott, L. C. (1990b). A view from physics. In M. Gardner et al. (Eds.), Toward a scientific practice of science education (pp. 3-30). Hillsdale, NJ: Erlbaum.

McDermott, L. C., & The Physics Education Group at the University of Washington. (1996). Physics by Inquiry (Volumes I & II). New York: Wiley.

McPeck, J. (1981). Critical thinking and education. New York: St Martin's.

Merriam, S. B. (1988). Case study research in education: A qualitative approach. San Francisco: Jossey-Bass.

Millar, R., & Driver, R. (1987). Beyond process. Studies in Science Education, 14, 33-62.

Mokros, J. & Russell, S. (1992). Children's concepts of average and representativeness. Working Paper 4-92. Cambridge, MA: TERC.

National Research Council. (1996). National science education standards. Washington, DC: National Academy Press.

National Research Council. (2000). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

National Science Foundation. (1996). Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology. Washington DC: National Science Foundation.

Nersessian, N. J. (1989). Conceptual change in science and in science education. Synthese, 80, 163-183.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. International Journal of Science Education, 21, 553-576.

Norris, S. P. (1985). The choice of standard conditions in defining critical thinking competence. Educational Theory, 35 (1), 97-107.

Norris, S. P., & Phillips, L. M. (1994). Interpreting pragmatic meaning when reading popular reports of science. Journal of Research in Science Teaching, 31, 947-967.

Nott, M., & Wellington, J. (1995a). Critical Incidents in Science. Milton Keynes: Open University Press.

Nott, M., & Wellington, J. (1995b). Critical incidents in the science classroom and the nature of science. School Science Review, 76 (276), 41-46.

Osborne, R., & Freyberg, P. (1985). Learning in science. Birkenhead, Auckland: Heinemann.

Palmer, D. (1997). The effect of context on students' reasoning about forces. International Journal of Science Education, 19, 681-696.

Patton, M. Q. (1980). Qualitative research methods. Beverly Hills, CA: Sage.

Patton, M. Q. (1990). Qualitative Evaluation and Research Methods. London: Sage.

Pea, R., & Greeno, J. G. (1990). Reflections on directions of reform in mathematics education. A paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Posner, G. J., & Gertzog, W. A. (1982). The clinical interview and the measurement of conceptual change. Science Education, 66, 195-209.

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

Pressley, M., Snyder, B. L., & Cariglia-Bull, T. (1987). How can good strategy use be taught to children: Evaluation of six alternative approaches. In S. M. Cormier & J. D. Hagman (Eds.), Transfer of learning (pp. 81-120). New York: Academic.

QSR NUDIST^tm [Computer software]. (1993). Victoria, Australia: Qualitative Solutions and Research Pty Ltd.

Renner, J. & Lawson, A. (1973). Promoting intellectual development through science teaching. Physics Teacher, 11, 273-275.

Rubba, P. & Anderson, H. (1978). Development of an instrument to assess secondary school students' understanding of the nature of scientific knowledge. Science Education, 62, 449-458.

Ruiz-Primo, M. A., Baxter, G. P., & Shavelson, R. J. (1993). On the stability of performance assessments. Journal of Educational Measurement, 30 (1), 41-53.

Salomon, G., & Globerson, T. (1987, October). Rocky roads to transfer. The second annual report to the Spencer Foundation. Israel: Tel-Aviv University.

Sandoval, W. A., & Reiser, B. J. (1997). Exploring explanations in high school biology. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. Developmental Psychology, 32, 102-119.

Schauble, L., Glaser, R., Duschl, R., Schulze, S., & John, J. (1994). Experimentation in the science classroom. Journal of the Learning Sciences, 4, 131-166.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. Journal of Research in Science Teaching, 28, 859-882.

Schoenfeld, A. H. (1985). Making sense of "out loud" problem-solving protocols. The Journal of Mathematical Behavior, 4, 171-191.

Schwab, J.J. (1964). Structure of the disciplines: Meanings and significances. In G. W. Ford & L. Pugno (Eds.), The structure of knowledge and the curriculum (pp. 6-30). Chicago: Rand McNally.

Schwab, J. J. (1966). The teaching of science as enquiry. Cambridge, MA: Harvard University Press.

Schwab, J. J. (1978). Education and the structure of the disciplines. In I. Westbury & N. J. Wilk (Eds.), Science, curriculum and liberal education (pp. 229-276). Chicago: University of Chicago Press.

Sere, M. G., Journeaux, R., & Larcher, C. (1993). Learning the statistical analysis of measurement errors. International Journal of Science Education, 15, 427-438.

Shayer, M., & Adey, P. (1981). Towards a science of science teaching. London: Heinemann.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15 (2), 4-14.

Smith, D., Conway, P., & Levine-Rose, S. (1995). The re-construction of preservice elementary teachers' conceptions of the nature of science. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California.

Smith, D. C., & Neale, D. C. (1989). The construction of subject matter knowledge in primary science teaching. Teaching and Teacher Education, 5 (1), 1-20.

Smith, E. L. (1987). What besides conceptions needs to change in conceptual change learning? In J. D. Novak (Ed.), Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics (pp. 424-433). Ithaca, NY: Department of Education, Cornell University.

Solano-Flores, G., Jovanovic, J., Shavelson, R. J., & Bachman, M. (1999). On the development and evaluation of a shell for generating science performance assessments. International Journal of Science Education, 21, 293-315.

Solomon, J. (1991). Group discussions in the classroom. School Science Review, 72, 29-34.

Sparks, D. (1997). Reforming teaching and reforming staff development: An interview with Susan Loucks-Horsley. Journal of Staff Development, 18 (4), 20-23.

Stake, R. E. (1995). The art of case study research. London: Sage.

Stone, C. A. & Day, M. C. (1978). Levels of availability of a formal operational strategy. Child Development, 49, 1054-1065.

Strang, J. (1990). Measurement in school science. Assessment Matters No. 2. London: SEAC/EMU.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin, & Y. S. Lincoln (Eds.), Handbook of Qualitative Research (pp. 273-285). Thousand Oaks, CA: Sage.

Suchman, R. (1966). Developing inquiry. Chicago: Science Research Associates.

Tabak, I., & Reiser, B. J. (1999). Steering the course of dialogue in inquiry-based science classrooms. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Tamir, P. (1983). Inquiry and the science teacher. Science Education, 67, 657-672.

Tamir, P. (1989). Training teachers to teach effectively in the laboratory. Science Education, 73, 59-69.

Tobias, S. (1992). Revitalizing undergraduate science: Why some things work and most don't. Tucson, AZ: Research Corporation.

Trumper, R. (1996). A cross-college age study about physics students' conceptions of force in pre-service training for high school teachers. Physics Education, 31, 227-236.

van Tilburg, P. A., Verloop, N., & Vermunt, J. D. (1999). Changes in teachers' knowledge and beliefs about the teaching of inquiry skills. A paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.

Varelas, M. (1997). Third and fourth graders' conceptions of repeated trials and best representatives in science experiments. Journal of Research in Science Teaching, 9, 853-872.

White, R., & Gunstone, R. (1992). Probing understanding. London: The Falmer Press.

Wiersma, W. (1995). Research methods in education: An introduction. Boston: Allyn and Bacon.

Wollman, W. (1977). Controlling variables: Assessing levels of understanding. Science Education, 61, 371-383.

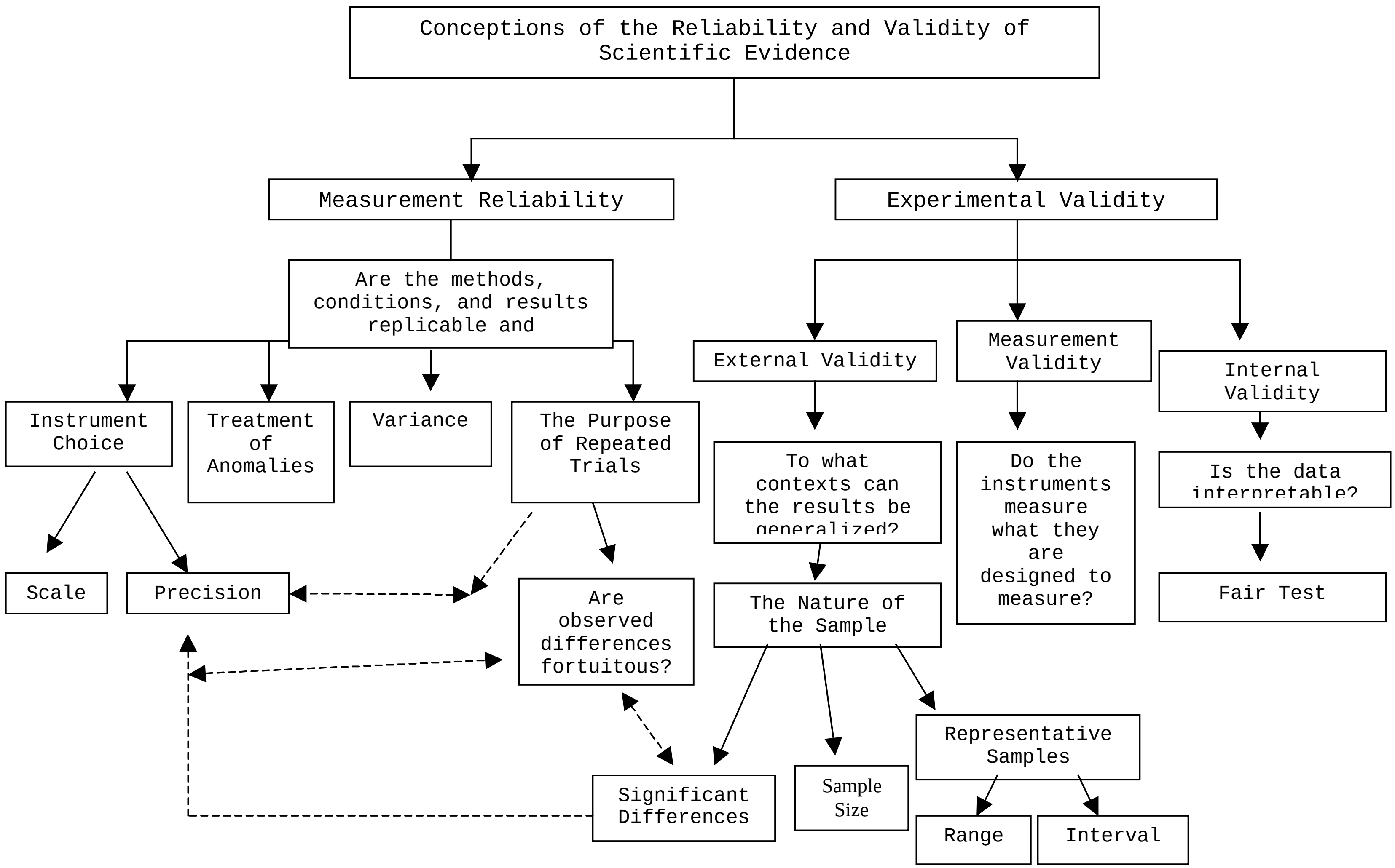
Wollman, W. & Lawson, A. (1977). Teaching the procedure of controlled experimentation: A Piagetian approach. Science Education, 61, 57-70.

Woolnough, B., & Allsop, T. (1985). Practical work in science. Cambridge, UK: Cambridge University Press.

Yager, R. E. (1991). The centrality of practical work in the science/technology/society movement. In B. Woolnough (Ed.), Practical Science (pp. 21-30). Milton Keynes: Open University Press.

Yin, R. K. (1994). Case study research: Design and methods. London: Sage.

APPENDIX A: Conceptions of the Reliability and Validity of Scientific Evidence



APPENDIX B
INFORMED CONSENT FORM

INFORMED CONSENT FORM FOR BEHAVIORAL RESEARCH STUDY

The Pennsylvania State University

Title of Project: Investigating conceptions of scientific evidence
Persons in Charge:

Joseph A. Taylor)	Thomas M. Dana
(Graduate Assistant)	(Associate Professor)
113 Chambers Bldg.	167 Chambers Bldg.
jat178@psu.edu	tdana@psu.edu
(814) 865-1544	(814) 865-6565

- **This section provides and explanation of the study in which you will be participating:**
- The study in which you will participate is part of research intended to explore the conceptions of scientific evidence held by pre-service secondary science education majors, practicing secondary science teachers, and non-teaching science majors.

- This study will involve the video recording of four interviews, including the completion of two questionnaires. Participation in this study will require about 4 hours of your time. The tapes will be coded to protect your confidentiality and stored in a secure (combination locked) location. Only the person in charge will have access to the tapes and to any identifying codes.

- **This section describes your rights as a research participant:**
  - A. You may ask any questions about the research procedures and these questions will be answered. Further questions should be directed to Mr. Joseph A. Taylor or Dr. Thomas M. Dana.

  - B. Your participation in this research is confidential. Only the persons in charge will have access to your identity and to information that can be associated with your identity. In the event of publication or presentation of this research, no personally identifying information will be disclosed.

  - C. Your participation is voluntary. You are free to stop participating at any time or to decline to answer any specific questions without penalty.

- **This section indicates that you are giving your informed consent to participate in the research:**

Participant:

- I agree to participate in a systematic investigation of conceptions of scientific evidence as an authorized part of the education and research program of The Pennsylvania State University.
- I understand the information given to me and have received answers to any questions I may have had about the research procedure. I understand and agree to the conditions of this study as described.
- To the best of my knowledge, I have no physical or mental illnesses/difficulties that would increase the risk to me by participating in this study.
- I understand that my participation in this research is voluntary, and that I may withdraw from the study at any time by notifying the person in charge.
- I understand that my participation in or withdrawal from this study will not affect my standing in any course at Penn State University.
- I understand that I will receive a signed copy of this consent form.
- I understand that all videotapes and audiotapes will be destroyed no later than May 2002.

____ _______________

Signature Date

Researcher:

I certify that the informed consent procedure has been followed, and that I have answered any questions from the participant above as fully as possible.

____ _______________

Signature Date

APPENDIX C

THINK ALOUD EXPERIMENTAL DESIGN INTERVIEW: THE RESISTANCE OF A WIRE

Investigating the resistance of a wire

1. What variables do you think affect the amount of resistance provided by a cylindrical conductor (wire)?

2. How does each variable affect resistance?

I would like you to design an experiment that would allow you to determine the relationship between the length of a wire and the wire's resistance. After you are finished, I will be asking you to describe how you plan to collect, organize, and represent your data, as well as why do you prefer to use these methods. Please describe your procedure thoroughly enough that a novice physics student could replicate your experiment.

Follow-up Prompts:

3. Why do you prefer this ammeter?
4. How do you plan to connect your ammeter? Why?
5. Why do you prefer to vary the length of the wire this way?
6. Why do you prefer "x" number of trials?
7. Is this a fair test? Why?

Imagine that your students conducted the experiment that you have described and collected the following data…

(Supply Appropriate Data Tables and Graphs Here)

The data tables provided will be designed to determine whether the subject addresses validity and reliability issues such as:

8. Treatment of Anomalies
9. Conclusions based upon small or fortuitous differences.
10. Variance
11. Range and Interval
12. Generalization to other contexts

APPENDIX D

ANALYSIS OF CLASSROOM PASSAGES: THE RESISTANCE OF A WIRE

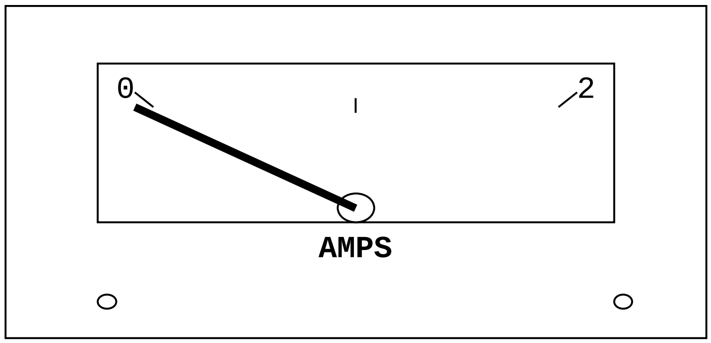
The following set of questions asks you to place yourself into a hypothetical teaching situation. In this teaching scenario, you are a high school physics teacher and have required your students (in small groups) to determine the factors that affect the resistance of a conducting wire. In addition to simply identifying the factors, you have asked your students to describe the nature of the relationship between each factor and the resistance. Each student group was required to present their findings to the class. In this presentation, you requested that compelling evidence be used to support all claims.

1. What factors would you expect the students to identify?

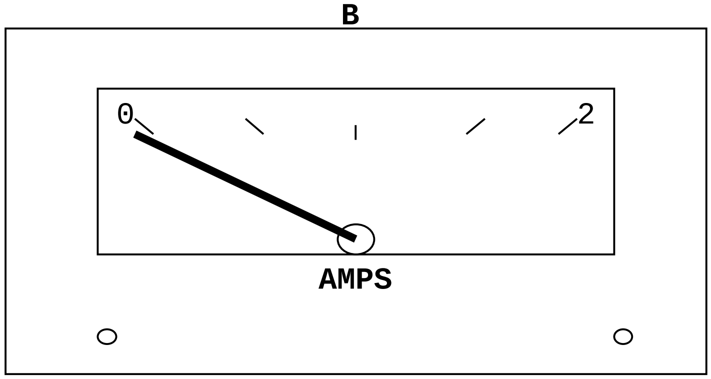
2. Briefly, describe what you expect the students to find out about the nature of each relationship.

3. As the students prepared to begin the investigation, one of the student groups indicated that they planned to use an ammeter and Ohm's law to determine the resistance of the wire. Since a variety of ammeters were available, the students asked you which one would be appropriate. Which of the ammeters below would you recommend if you wanted your students to collect the most reliable data possible? Explain your reasoning.

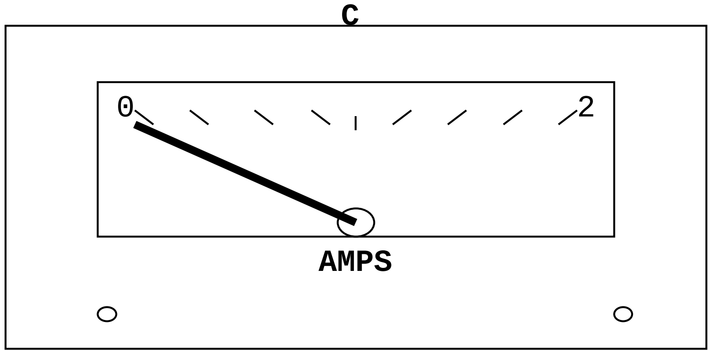
A



B



C



a. Ammeter A b. Ammeter B c. Ammeter C

4. One of the student groups claimed that increasing the length of the wire caused the resistance to decrease. The students supported this claim with following data table.

Length of Wire	Resistance
Tungsten-100cm	200 ohms
Aluminum-105cm	175 ohms
Copper-110cm	165 ohms
Silver-115cm	140ohms

What is your evaluation of this claim?

5. Three of the student groups claimed that the length of the wire affected the resistance. The data tables constructed by these student groups are shown below.

Group A

Length of the wire	5cm	65cm	70cm	80cm	85cm
Resistance	0.00025 ohms	0.0032 ohms	0.0035 ohms	0.0040 ohms	0.0043 ohms

Group B

Length of the wire	18cm	36cm	54cm	72cm	90cm
Resistance	0.0009 ohms	0.0018 ohms	0.0027 ohms	0.0036 ohms	0.0045 ohms

Group C

Length of the wire	1cm	2cm	3cm	4cm	5cm
Resistance	0.00005 ohms	0.0001 ohms	0.00015 ohms	0.0002 ohms	0.00025 ohms

How would you evaluate these claims?

Which set(s) of data are most likely to fully illustrate the relationship between the length of the wire and its resistance? Why?

Do you have concerns with the other data set(s)? If so, explain your concerns.

6. During the course of the investigation, Group B suspected that the ammeter they used to make current measurements was not working properly. However, they did not question the average of their trials because it matched the average obtained by Group A and compared well with the theoretical prediction of two amps (see Data Tables below).

Group D

Trial	1	2	3	4	5	Avg. Current Measured
Cross-sectional Area of Wire	3.4×10^{-6} m^2	3.4×10^{-6} m^2	3.4×10^{-6} m^2	3.4×10^{-6} m^2	3.4×10^{-6} m^2	
Length of Wire	50cm	50cm	50cm	50cm	50cm	
Current	1.9A	2.0A	1.8A	2.2A	2.1A	2.0A

Group E

Trial	1	2	3	4	5	Avg. Current Measured
Cross-sectional Area of Wire	3.4×10^{-6} m^2	3.4×10^{-6} m^2	3.4×10^{-6} m^2	3.4×10^{-6} m^2	3.4×10^{-6} m^2	
Length of Wire	50cm	50cm	50cm	50cm	50cm	
Current	1.0A	3.0A	0.5A	3.5A	2.0A	2.0A

What do you think of Group E’s experiment? Explain.

7. In their presentation, one student group claimed that the resistance of a wire increased when the length of the wire in the circuit was increased. This group noted, however, that the resistance did not increase uniformly with uniform increases in the wire's length. This group based this conclusion upon the following data that were collected with a sensitive multimeter.

Length of wire	20cm	40cm	60cm	80cm	100cm	120cm	140cm
Trial 1	0.06Ω	0.11Ω	0.19Ω	0.24Ω	0.31Ω	0.36Ω	0.43Ω
Trial 2	0.05Ω	0.12Ω	0.18Ω	0.24Ω	0.31Ω	0.35Ω	0.43Ω
Trial 3	0.06Ω	0.13Ω	0.18Ω	0.24Ω	0.29Ω	0.36Ω	0.42Ω
Trial 4	0.06Ω	0.13Ω	0.18Ω	0.26Ω	0.30Ω	0.37Ω	0.43Ω
Trial 5	0.07Ω	0.26Ω	0.17Ω	0.22Ω	0.29Ω	0.36Ω	0.44Ω
Avg. Resistance Measured	0.06Ω	0.15Ω	0.18Ω	0.24Ω	0.30Ω	0.36Ω	0.43Ω

Do you agree with this conclusion? Why or Why not?

If you do not agree, what would you recommend to this group? Why?

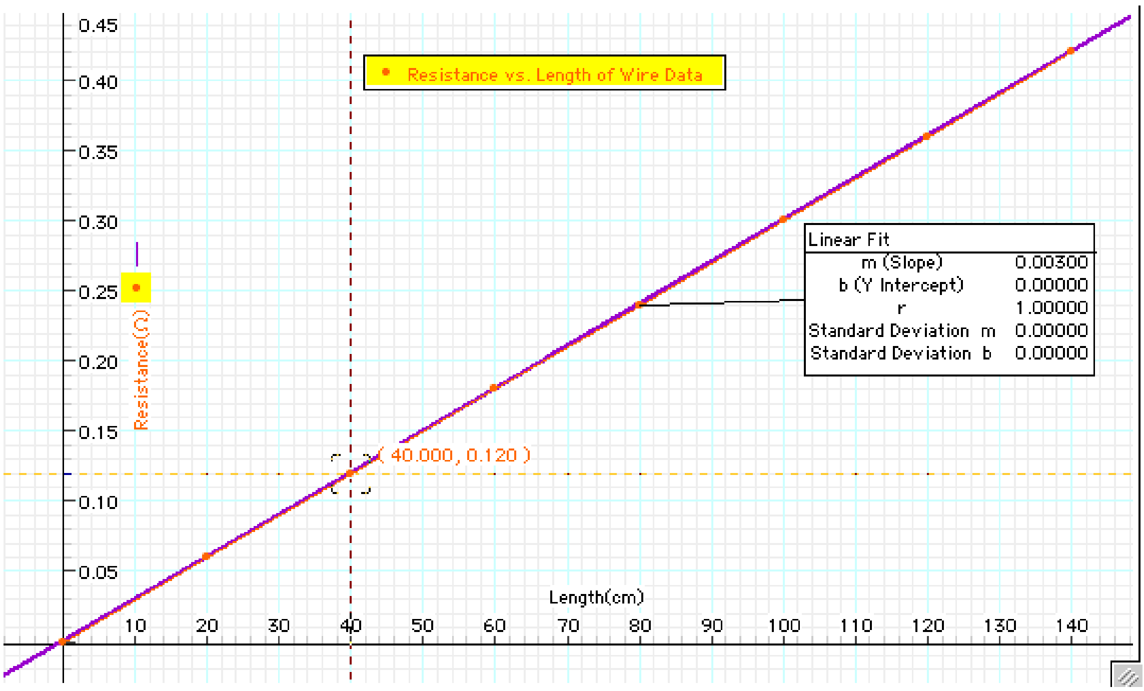
8. In one presentation, a student group thoroughly described their experimental procedure. The students mentioned that they constructed a circuit consisting of a high voltage power supply, some copper wire, and an ammeter. The ammeter was used to measure the current in the circuit and then Ohm's law (V = IR) was used to determine the resistance of the wire. The students also described how, in some of the trials, they used the high voltage power supply to produce a current of five amps in the wire.

If at all, how would you respond to this group's experimental procedure?

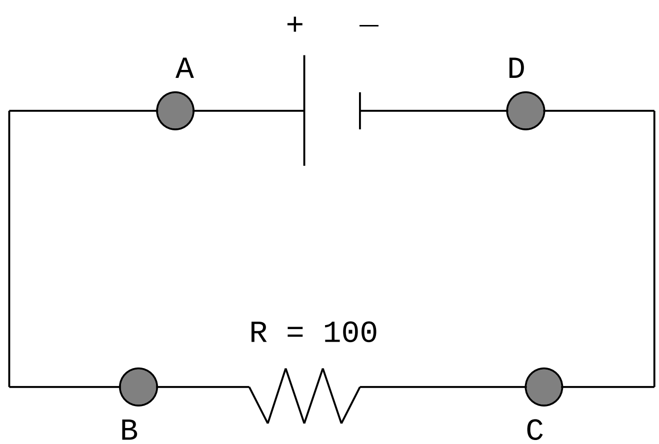
9. Referring to the graph below, a student group concluded…

“In sum, the resistance of a wire will increase by 0.003Ω every time its length is increased by 1cm.”

How would you respond to this claim? Why?



10. One of the student groups suspected that the amount of current in a series circuit depended upon the location in the circuit at which the current was measured. The students investigated this hypothesis by constructing the circuit diagram shown below and varying the placement of the ammeter each time. The ammeter was placed at locations: A, B, C, and D. The circuit was not changed in any other way. A current measurement was taken when the ammeter was placed at each of the locations.



The students supported this conclusion with the data provided in the table below.

Location	Current
A	1.25A
B	1.24A
C	1.23A
D	1.23A

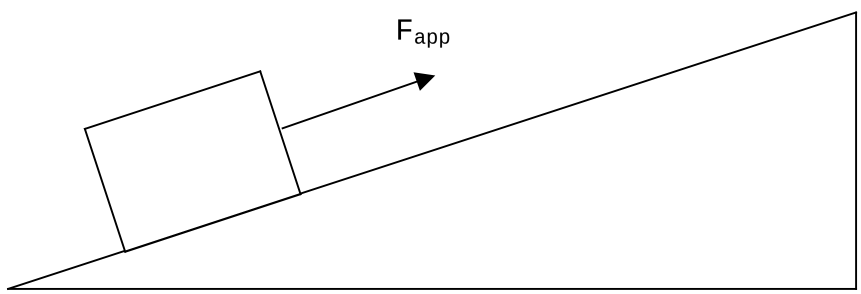
How would you respond to this group’s evidence? Explain.

APPENDIX E

ANALYSIS OF CLASSROOM PASSAGES: THE INCLINED PLANE

The following set of questions asks you to place yourself into a hypothetical teaching situation. In this teaching scenario, you are a high school physics teacher and have required your students (in small groups) to determine the factors that affect the minimum applied force (F_{app}) necessary to pull a wooden block up an inclined plane (see Figure 1). In addition to simply identifying the factors, you have asked your students to describe the nature of the relationship between each factor and the necessary force. Each student group was required to present their findings to the class. In this presentation, you requested that compelling evidence be used to support all claims.

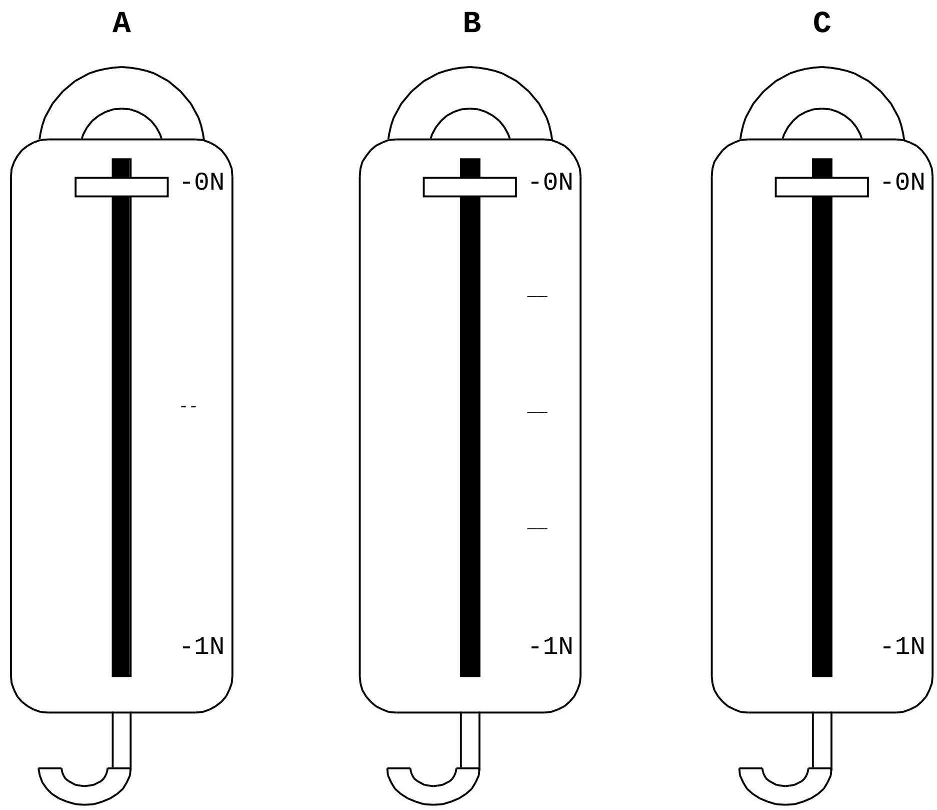
<u>Figure 1</u>.



1. What factors would you expect the students to identify?

2. Briefly, describe what you expect the students to find out about the nature of each relationship.

3. As the students prepared to begin the investigation, one of the student groups indicated that they planned to use a spring balance to measure the applied force. Since a variety of spring scales were available, the students asked you which one would be appropriate. Which of the spring scales below would you recommend if you wanted your students to collect the most reliable data possible? Explain your reasoning.



a. Spring Scale A

b. Spring Scale B

c. Spring Scale C

4. One of the student groups claimed that increasing the mass of the block caused the required force to decrease. The students supported this claim with following data table.

Mass of Block	Average Applied Force Necessary	Slope of Incline
100g	1.05N	60˚
105g	1.00N	40˚
110g	0.80N	20˚
115g	0.65N	10˚

What is your evaluation of this claim?

5. Three of the student groups claimed that the slope of the incline affected the necessary applied force. The data tables constructed by these student groups are shown below.

Group A

Slope of Incline	5°	65°	70°	80°	85°
Required Force	4.8N	10.6N	10.5N	10.3N	10.1N

Group B

Slope of Incline	18°	36°	54°	72°	90°
Required Force	6.8N	8.9N	10.2N	10.5N	9.8N

Group C

Slope of Incline	1°	2°	3°	4°	5°
Required Force	4.1N	4.3N	4.4N	4.6N	4.8N

How would you evaluate these claims?

Which set(s) of data are most likely to fully illustrate the relationship between the force necessary to pull an object up an inclined plane and the slope of the inclined plane? Why?

Do you have concerns with the other data set(s)? If so, explain your concerns.

6. During the course of the investigation, Group E suspected that the spring scale they used to make force measurements was not working properly. However, they did not question the average of their trials because it matched the average obtained by Group D and compared well with the theoretical prediction of seven Newtons (see Data Tables below).

Group D

Trial	1	2	3	4	5	Avg. Applied Force Required
Mass of Block	500g	500g	500g	500g	500g	
Angle of incline	40˚	40˚	40˚	40˚	40˚	
Applied Force Required	6.5N	7.0N	7.5N	7.0N	7.0N	7N

Group E

Trial	1	2	3	4	5	Avg. Applied Force Required
Mass of Block	500g	500g	500g	500g	500g	
Angle of incline	40˚	40˚	40˚	40˚	40˚	
Applied Force Required	10.5 N	4.5N	9.5N	5.5N	5.0N	7N

What do you think of Group E’s experiment? Explain.

7. In their presentation, one student group claimed that required applied force increased when the mass of the block was increased. This group noted, however, that the required force did not increase uniformly with uniform increases in the block's mass. This group based this conclusion upon the following data.

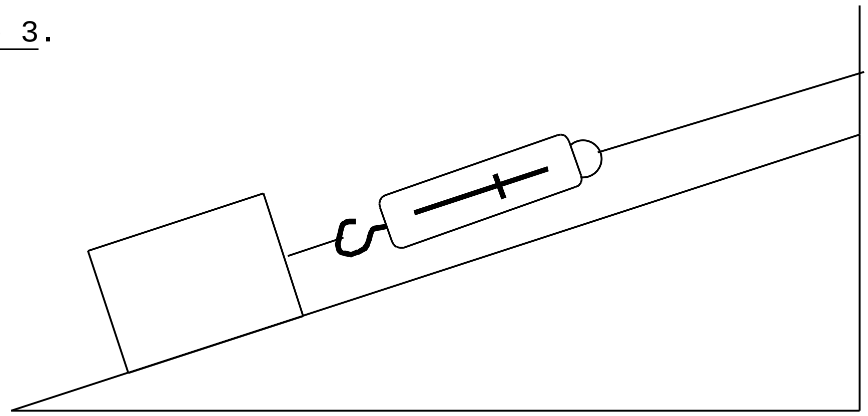
Mass of the Block	100g	200g	300g	400g	500g	600g	700g
Trial 1	0.80N	1.70N	2.40N	3.20N	4.20N	4.80N	5.50N
Trial 2	0.80N	1.50N	2.40N	3.15N	4.10N	4.75N	5.50N
Trial 3	0.85N	1.65N	2.40N	3.25N	3.90N	4.80N	5.60N
Trial 4	0.75N	1.55N	2.35N	3.20N	3.80N	4.80N	5.70N
Trial 5	0.80N	1.60N	2.45N	0.70N	4.00N	4.85N	5.70N
Avg. Applied Force Necessary	0.80N	1.60N	2.40N	2.70N	4.00N	4.80N	5.60N

Do you agree with this conclusion? Why or Why not?

If you do not agree, what would you recommend to this group? Why?

8. In one presentation, a student group thoroughly described their experimental procedure. The students mentioned that the minimum applied force could be measured by attaching (using string) the end of the spring scale to a post at the top of the incline (see Figure 3). The students argued that measuring the applied force this way would account for the block's weight, the slope of the incline, and any frictional forces.

Figure 3.

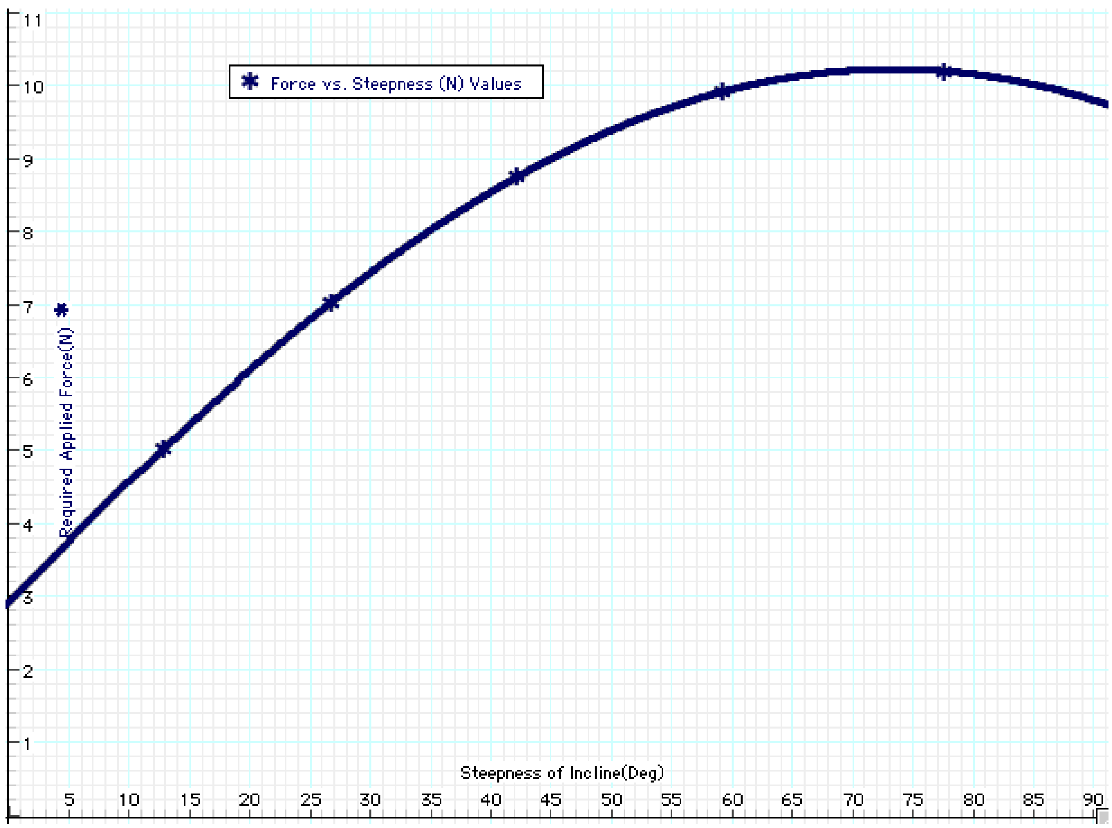


If at all, how would your respond to this group?

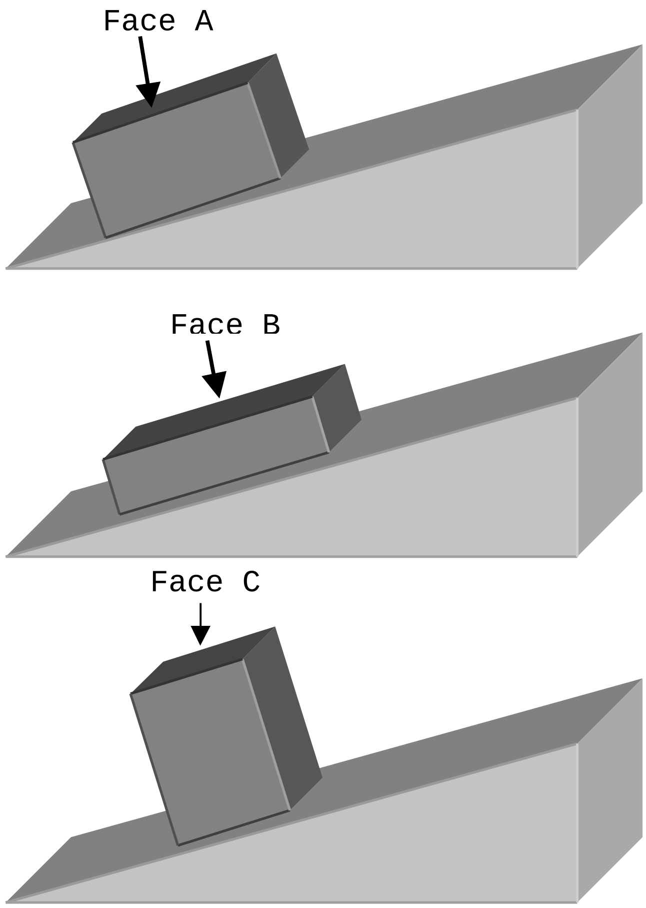
9. Referring to the graph provided, a student group concluded…

“In sum, the applied force necessary to pull an object up an inclined plane will typically reach a maximum when the angle of incline is between 70 and 75 degrees.”

How would you respond to this claim? Why?



10. One of the student groups suspected that the minimum applied force necessary to overcome friction depended on how the block was placed on the incline. That is, the minimum applied force needed to initiate motion up the incline would vary when the block was placed on each of its three, different sized faces (see figure below).



Using the spring scale provided, the students pulled just hard enough on the block to initiate motion. This was done: once on Face A, once on Face B, and once on Face C. The students concluded that the minimum applied force necessary to initiate motion depended on the size (area) of the face that the block was dragged upon. The students supported this conclusion with the data provided in the table below.

Face	Required Applied Force
A	9.70N
B	9.65N
C	9.75N

How would you respond to this group's evidence? Explain.

VITA

Joseph A. Taylor
528 Galen Dr.
State College, PA 16803
(814) 234-0910/jat178@psu.edu

EDUCATION

Doctor of Philosophy, Curriculum and Instruction (Science Education), August 2001 (anticipated). The Pennsylvania State University, University Park, Pennsylvania

Master of Secondary Mathematics Education. December 1992
Ohio University, Athens, Ohio

Bachelor of Science in Physics Education. December 1991
California University of Pennsylvania, California, Pennsylvania

TEACHING and PROFESSIONAL EXPERIENCE

Graduate Teaching Assistant
1998 - Present: The Pennsylvania State University, University Park, Pennsylvania
Responsibilities include:
-Teaching of ENGR 497F: *Fundamentals of Science, Technology, and Engineering Design* – a physical science course designed for prospective elementary school teachers.
-Teaching and supervision of prospective secondary science teachers enrolled in SCIED 412 (*Teaching of Science in the Secondary School II*).

PASCO Technology Educator
July 2000 to present: PASCO Scientific, Roseville, California.
Responsibilities include:
-Planning and conducting workshops for middle school and high school science teachers which focus on computer -based collection and analysis of scientific data.

Science and Mathematics Teacher
August 1993 to June 1998: Middletown Area High School, Middletown, Pennsylvania.
Responsibilities included:
-Teaching of *Physics I, Physics II*, and *Algebra II*.

PUBLICATIONS

Taylor, J. A. (2000). Preparing for Pre-service Teacher Supervision: Developing a Supervisory Model. Journal of Critical Inquiry into Curriculum and Instruction, 2 (2), 38-40.

Taylor, J. A. (2001). Using a practical context to encourage conceptual change: An instructional sequence in bicycle science. *School Science and Mathematics*, 101, 117-124.

Taylor, J. A., Dana, T. M., & Tasar, M. F. (2001). An Integration of Simple Materials and Complex Ideas: Description of an Instructional Sequence in Statics. *International Journal of Engineering Education* 17, (pp. TBD).

Taylor, J. A., Lunetta, V. N., Dana, T. M., & Tasar, M. F. Bridging Science and Engineering: An Integrated Course for Non-Science Majors. Accepted for publication in the *Journal of College Science Teaching* – November 2000.

Taylor, J.A. & Dana, T. M. (in press). Trouble with Trusses. In D. Tippins, T. Koballa, & B. Payne (Eds.), *Science Teaching and Learning in the Elementary Classroom: A Classroom Case Handbook*. Allyn & Bacon.

PROFESSIONAL ASSOCIATIONS

American Association of Physics Teachers (A.A.P.T.)
National Association of Research in Science Teaching (N.A.R.S.T)
National Science Teachers Association (N.S.T.A.)
American Society of Engineering Educators (A.S.E.E.)
School Science and Mathematics Association (S.S.M.A.)