JOURNAL OF PHYSICS TEACHER EDUCATION ONLINE

Vol. 1, No. 4 www.phy.ilstu.edu/jpteo March 2003

JPTEO

INSIDE THIS ISSUE

Carl J. Wenning
JPTEO EDITOR-IN-CHIEF
Department of Physics
Illinois State University
Campus Box 4560
Normal, IL 61790-4560
wenning@phy.ilstu.edu

Improving Physics Teacher Preparation

During the past nine years I have dedicated a vast amount of time, effort, and resources to improving the preparation of physics teacher candidates here at Illinois State University. A large part of the improvement process has been based on ideas generated through personal professional development. Another part has been based upon aligning the program with Standards for Science Teacher Preparation promulgated by the National Science Teachers Association. As the result of recent compliance work related to a program rejoinder, this issue of *JPTEO* is being published later than I had originally intended. Nonetheless, the delay is well worth it if physics teacher preparation will improve as a result.

As a physics teacher education program coordinator, many people with whom I have regular contact seem to feel that I should resent spending large amounts time and effort "jumping through hoops." Actually, I find the accreditation process to be a stimulating source of new ideas for improving ISU's physics teacher education program -- especially in the area of performance assessment. For instance, during the past two months I have created ten new performance-based summative assessments dealing with different knowledge and skill areas of teacher candidate performance. Eight of these are preservice-level assessments associated with five of our six pedagogically-oriented physics courses. Knowledge and skills assessed run the range from knowing about the nature of science and demonstrating how to conduct scientific inquiry, to exhibiting skills of teaching and including the social context. The remaining two performance assessments are at the induction level, and deal with the student teaching practicum and creation of a professional teaching portfolio. All assessment instruments have requirements, rubrics, and specific criteria describing acceptable performance. All assessment instruments are standards based.

Last year I also included for the first time, and by way of personal professional development, the process known as Lesson Study (outlined in *The Teaching Gap* by Stigler and Hiebert). This in-depth lesson planning, teaching, evaluation, and revision process has been described by my students as "the best single thing I have done to date in preparation for teaching." I've heard this comment expressed in other ways more than once. This year I will give the process considerably more emphasis. If, as a physics teacher educator, you have never used the Lesson Study approach, I strongly suggest that you do so.

Each quarter our *JPTEO* authors take the time and effort to prepare quality articles that can be used to enhance physics teacher candidate preparation, or to improve the teaching skills of in-service teachers by way of professional development. This quarter's issue is no exception. Taking our authors' lead, I can't help but sharing what I have learned and developed through the NSTA accreditation process. Another part of the reason I am sharing these resources is because other science teacher educators have reviewed them and found them to be helpful. In the web site address provided below, readers can find assessment activities associated with the following ten areas:

Nature of Science Project	***Social Context Project***
Scientific Inquiry Project	***Physics Exam Project***
Building Bridges Project	***Lesson Study Project***
Unit Plan Project	***Professional Portfolio***
Curriculum Project	***Student Teaching***

Those who coordinate physics teacher education programs might want to examine these performance-based assessments. Even though the assessments have yet to receive the "blessing" of NSTA reviewers, I want to share the project requirements and their associated rubrics with others who might find themselves in a similar situation. Even if they aren't perfect, they can serve as a basis of refinement work. That's what action research is all about. I hope that our readers will share some of their "work in progress." You may examine mine at the following URL: *http://www.phy.ilstu.edu/ptefiles/summative.html.*

Carl J. Wenning
JPTEO EDITOR-IN-CHIEF

JOURNAL OF PHYSICS TEACHER EDUCATION ONLINE

Journal of Physics Teacher Education Online is published by the Department of Physics at Illinois State University in Normal, Illinois. Editorial comments and comments of other authors do not necessarily reflect the views of Illinois State University, the Department of Physics, or its editor-in-chief. *JPTEO* is available online at *www.phy.ilstu.edu/jpteo.* To subscribe to this journal, send an e-mail to the editor indicating that you wish to be added to the notification list. When issues are published online, subscribers will receive electronic notification of availability. *JPTEO* is published on a quarterly basis. Issues appear during March, June, September, and December. It is available free of charge through the Journal's web site. *JPTEO* is downloadable in portable document file (PDF) format.

Creating and maintaining any sort of journal requires a commitment from its readership to submit articles of interest and worth in a timely fashion. Without such contributions, any journal is bound to fail. It is hoped that *JPTEO* becomes a forum of lively exchange. It will become so only to the extent that its readers will submitting articles for consideration and publication. Detailed information about contributing to *JPTEO* can be found on the journal's web site.

An illustration of the complex nature of subject matter knowledge: A case study of secondary school physics teachers' evaluation of scientific evidence

Joseph A. Taylor
Professional Development Associate
The SCI Center at BSCS
5415 Mark Dabling Blvd.
Colorado Springs, CO 80918
jtaylor@bscs.org
719-531-5550

Thomas M. Dana
Hermanowicz Professor of Teacher Education
167 Chambers Bldg.
The Pennsylvania State University
University Park, PA 16802
tdana@psu.edu
814-865-6565

The principal aim of this study is to examine one prospective and two practicing physics teachers' evaluations of scientific evidence. The data from this study suggest that the three teachers frequently used their conceptions of scientific evidence in conjunction with physics subject matter conceptions while evaluating evidence. The data also indicate that the use of subject matter knowledge in conjunction with conceptions of evidence was more pronounced when evaluating certain types of scientific data and conclusions. Implications for physics teacher education suggest that physics and physics methods courses encourage teachers to conduct original research and to construct and present evidence-based arguments from this research for peer review and critique.

For the past several decades, science educators have recommended scientific literacy be improved (e.g., Bauer, 1992; Dewey, 1933; Herron, 1971; Kyle, 1980; Yager, 1991). These recommendations highlighted 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 physics teacher educators. What types of understandings should physics 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 physics teachers must also 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 physics 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 (Taylor, 2001) 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 I). 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 subunits, specific conceptions of scientific evidence (e.g., the rationale for repeating trials) were treated as subunits 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 hypothesized to exist across this span because it was assumed that experience in struggling with student-generated data and conclusions might promote the development of certain conceptions of scientific evidence. Therefore, the differences in teaching experience could result in data that would allow the authors to construct a more thorough description of the case.

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.

The Protocols

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* (sample items to follow), 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 physics subject matter concepts with selected conceptions of scientific evidence when evaluating data and/or claims. That is, each participant used physics subject matter concepts in conjunction with selected conceptions of scientific evidence when evaluating data and conclusions. In addition, the authors observed that some conceptions of scientific evidence (e.g., measurement validity, statistical significance) tended to be integrated with physics subject matter concepts more extensively than other conceptions of scientific evidence (see Table I). These findings 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' physics subject matter concepts interacted with their conceptions of scientific evidence and describe how certain conceptions of scientific evidence seemed to be more extensively integrated with physics subject matter concepts than others (see Table I). In the following section, selected cross-participant analyses are provided as examples of differing degrees of integration with physics subject matter 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

Table I. *Conceptions of scientific evidence: Integration with physics subject matter concepts.*

Extensive Integration of Physics Subject Matter 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 II). 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 handwritten outline of their preferred experimental design.

After each participant described his or her experimental design, the authors asked him or her 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.

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.

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.

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' handwritten 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 II).

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 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 Physics Subject Matter Concepts - Example 2: *Measurement Validity*

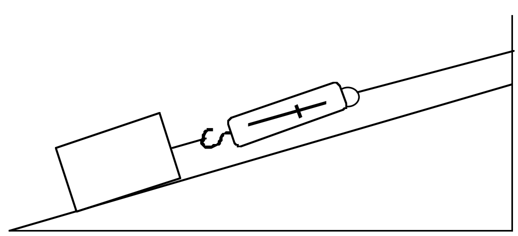
In some of the hypothetical scenarios presented in the *Analysis of Classroom Passages* protocol, the participants were presented with inappropriate conclusions based upon misuse of instruments (see Figure I).

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, Material	Length
Kurt	Length, Material, Temperature	Length, Material
Nina	Length, Voltage, Current	Voltage, Current

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

The participants tended to offer similar responses to the prompt shown in figure 1. That is, each participant was concerned that using the spring balance in this way might not be appropriate.

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?

Figure I. *Research prompt: Inclined Plane Classroom Passage.*

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.
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.

Nina:
N: Kinetic frictional forces could not be measured because the block is not moving.

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 Physics Subject Matter 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 II.

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.

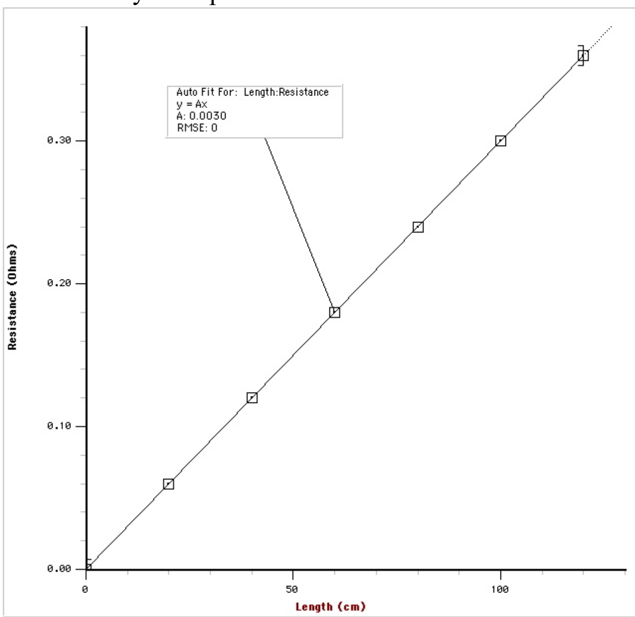
Betty (from the electricity-based prompt):
B: I would agree to this claim because that is what their graph is telling them. The slope of the graph is .003... however they need to clarify their claim because the slope of their graph is .003 but that does not mean the slope of a graph referring to other types of wire will be .003. When they make their claim they need to be specific to their experiment.

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 (emphasis added).

Referring to the graph below, a student group concluded…

"In sum, the resistance of a wire will increase by 0.003 _ every time it's length is increased by 1 cm."

How would you respond to this claim?



Referring to the graph below, 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 the inclined plane is between 70 and 75 degrees."

How would you respond to this claim?

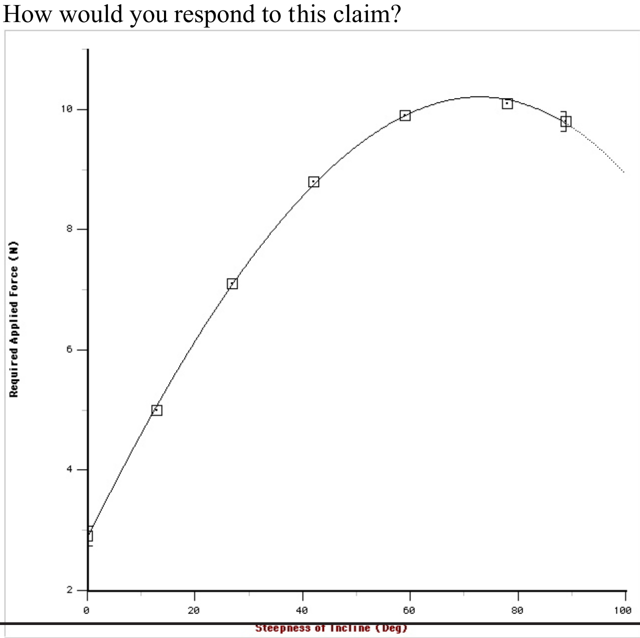


Figure II. *Research prompts regarding the Generalization of Conclusions.*

Kurt (from the inclined plane-based prompt):

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.

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.

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.

Nina (from the inclined plane-based prompt):

N: Well, it should continue to increase, shouldn't it? That's what I would think. So, I wouldn't agree with this conclusion.

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 of ρ/A (where ρ = the resistivity of the wire, and A = the cross-sectional area) is approximately equal to 0.003 Ω/cm].

The students' conclusion in the inclined plane-based prompt includes an over-generalization 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 preoccupation with the nonlinear nature of the graph probably drew their attention away from the applicability of the conclusion.

Limited Integration of Physics Subject Matter 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 III).

Group D - Inclined Plane Experiment

Trial	1	2	3	4	5	Avg. Applied Force Required
Mass of Block	900g	900g	900g	900g	900g	
Angle of incline	40°	40°	40°	40°	40°	
Applied Force Required	9.0N	10.0N	11.0N	10.0N	10.0N	10N

Group E – Inclined Plane Experiment

Trial	1	2	3	4	5	Avg. Applied Force Required
Mass of Block	900g	900g	900g	900g	900g	
Angle of incline	40°	40°	40°	40°	40°	
Applied Force Required	15.0N	7.0N	10.0N	13.0N	5.0N	10N

Group D – Resistance of a wire experiment

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

Group E – Resistance of a wire experiment

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

Figure III. *Research Prompts related to Reliability of Data.*

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 (from the electricity-based prompt):
I: Why do you think that group D's data is more reliable?
B: The data numbers found fell right around the desired average. I think consistent data is more reliable.

Betty (from the inclined plane-based prompt):
I: So, is group E's data 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.

Kurt (from the inclined plane-based prompt):
I: Why don't you think group E's data 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.

Nina (from the inclined plane-based prompt):

I: What do you think of Group E's 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.

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 understanding of the participants' physics subject matter was 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 in subject matter understanding, coupled with the consistency of responses to data reliability prompts, suggests that highly developed subject matter knowledge may not be *necessary* to think critically about the reliability of data.

The relative independence of conceptions of the reliability of data from physics subject matter 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 physics subject matter concepts. Betty's responses to prompts containing sweeping generalizations illustrated how a deeper understanding of physics subject matter 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 overgeneralizations in this context. 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 physics misconceptions. Specifically, Kurt and Nina's expectation of linearity in the applied force vs. angle of inclination graph 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 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 indicate that the participants integrated their physics subject matter concepts more often with certain conceptions of scientific evidence (control of variables, generalization of data, experimental validity) than with others (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 physics subject matter concepts more often than others.

It was also observed that the accuracy of certain physics subject matter concepts was influential to the participants' ability to identify contexts to which scientific conclusions might be applied. Future research might explore more deeply the nature of physics teachers' subject matter 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.

For Physics Teacher Education

The vision of physics education reform that has students actively engaged in investigating important questions, collecting data, making evidence-based claims, and arguing conclusions requires rather sophisticated subject matter knowledge for physics teachers. Schwab (1964), in theorizing about the different types of subject-matter expertise, distinguished between two types of subject matter knowledge, *substantive* knowledge and *syntactic* knowledge. Schwab viewed substantive knowledge as one's knowledge of the essential concepts, principles, and theories of the discipline. On the other hand, knowledge of the canons of evidence that guide inquiry in a discipline was referred to as syntactic knowledge. These canons include the norms for validating knowledge claims.

The integration of conceptions of scientific evidence with physics subject matter concepts observed in this study suggests that the critical evaluation of scientific evidence requires one to integrate both syntactic and substantive knowledge. This observation raises important issues for physics teacher education. How should physics teacher education programs address these knowledge bases so as to prepare teachers to evaluate student-generated scientific evidence? Demographic data provided by each participant indicated that each had taken several physics courses that "covered" some of the physics content discussed in this study. Therefore, the researchers suggest that the issue for teacher education program design may not be the number of physics courses taken but more so how the physics in these courses is learned.

In the past, some have tried to teach students critical thinking skills using a "general" approach (Ennis, 1989). 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 skills 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 skills are too content 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).

Since the data from this study indicated that the participants effectively integrated conceptions of scientific evidence with physics subject-matter concepts to evaluate scientific evidence, the authors recommend that courses in physics teacher education programs use instructional approaches that are likely to cultivate this effective partnership. One such approach would be to encourage practicing and prospective physics teachers to immerse themselves in a small set of semi-structured, scientific investigations. The data generated from these investigations could be woven into an evidence-based argument that is presented to peers for review and critique. This type of approach where students are encouraged to concurrently develop both substantive and syntactic concepts has been referred to as an *infusion* approach (Ennis, 1989).

The authors' recommendation for an infused approach was informed by the findings of this study as well as the recommendations of other scholars and researchers. For example, McDermott (1990) 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, among peers, 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 argument construction 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 with 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). These research findings also suggested that the ability to critically evaluate scientific evidence might be enhanced with practice.

Although the use of argument construction and peer review has received little attention in the teacher education literature, the lack of their use with teachers has been recognized (Smith, Conway, & Levine-Rose, 1995). The apparent disregard for argumentation in physics teacher education is of great concern since failure to engage students' in argumentation has been has been associated with the inability to critically evaluate scientific claims (Norris & Phillips, 1994; Solomon, 1991).

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

Since the ability to critically evaluate evidence might indeed be linked to practice, physics teacher educators could also provide practicing and prospective 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. Physics teacher educators could also provide practicing and prospective teachers with the opportunity to evaluate scientific evidence collected by actual secondary school students. To do this, physics teacher educators might call upon the resources and expertise of secondary school physics teachers from surrounding communities. Physics 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. In fact, this strategy has already been used successfully with practicing physics teachers. For example, Feldman (1993) found that the physics teachers involved in the *Physics Teacher Action Research Group* (PTARG) developed a deeper understanding of physics subject matter concepts as a result of their collaborative evaluation of student work.

We invite interested physics teacher eductors to contact us directly with reactions to this study as well as ideas about its implications for teacher education. Additional information about the larger study can be obtained by contacting the first author.

References:

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.

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.

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

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.

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.

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.

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.

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

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

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

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

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

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

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

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

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.

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

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

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.

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.

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.
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.
Linn, M. C., Clement, C., & Pulos, S. (1983). Is it formal if its not physics? *Journal of Research in Science Teaching, 20*, 755-770.
Linn, M. C., & Swiney, J. F. (1981). Individual differences in formal thought: Role of expectations and aptitudes. *Journal of Educational Psychology, 73*, 274-286.
Lubben, F., & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education, 18*, 955-968.
McDermott, L. C. (1990). 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.
McPeck, J. (1981). *Critical thinking and education*. New York: St Martin's.
National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
Nersessian, N. J. (1989). Conceptual change in science and in science education. *Synthese, 80*, 163-183.
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.
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.
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. (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.
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.
Solomon, J. (1991). Group discussions in the classroom. *School Science Review, 72*, 29-34.
Stake, R. E. (1995). *The art of case study research*. London: Sage.
Strang, J. (1990). *Measurement in school science*. Assessment Matters No. 2. London: SEAC/EMU.
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.
Taylor, J. A., & Dana, T. M. (2001). *Exploring secondary school physics teachers' conceptions of scientific evidence*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, St. Louis, Missouri.
Taylor, J. A. (2001). *Secondary school physics teachers' conceptions of scientific evidence: A collective case study*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park.

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.

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.

JPTEO

Implications of Modeling Method training on physics teacher development in California's Central Valley

David Andrews
Department of Biology
California State University, Fresno
davidan@csufresno.edu

Michael Oliver, FCEPT Evaluator
moliver@mail.telis.org

James Vesenka
Department of Chemistry and Physics
University of New England
jvesenka@une.edu

The Fresno Collaborative for the Excellence in the Preparation of Teachers is an NSF sponsored program aimed at recruiting and training outstanding math and science teachers for K-12 teaching positions in California's Central Valley. This multifaceted collaborative has held a summer physics modeling workshop for the past four years, in the same vein as Arizona State University's modeling instruction program. The results from two summers worth of interviews and assessments on modeling training indicate improved teacher comprehension of physics content as well as enthusiastic support from the participants in the modeling approach. Lastly, follow-up interviews with teachers have indicated that modeling instruction has greatly influenced their teaching styles and that they find their students more attentive and enthusiastic participants in their classrooms.

Background

The Fresno Collaborative for the Excellence in the Preparation of Teachers (FCEPT) is a partnership between California State University, Fresno, Fresno City College (FCC) and the Fresno Unified School District (FUSD), funded by a four-year grant from the NSF starting in 1999. **A recruitment pipeline has been established with surrounding community colleges. The experience and wisdom of other CETPs in California have been essential in the development of effective in-service and pre-service training activities.** We have received industrial sponsorship through a GTE technology grant over the summers of 2000-01 for our Project Science students. The support helped demonstrate ways to more effectively teach science through mathematically-based tools such as graphing calculators, sensors, and interfaces. Sixteen other local, state and federal collaborators have helped to provide teacher training in content and through experiential learning. These collaborations are designed to satisfy one or more of the following project goals.

1. Recruitment of committed and academically successful pre-service science and mathematics teachers into the teacher preparation program for middle and secondary school science and mathematics.
2. Pre-service mentoring of students by university/college district science and math and education faculty.
3. Early field experiences for pre-service undergraduates in science and mathematics.
4. Development of a five-year academic program combining pedagogy with content.
5. Revision and field-testing of university/college science and/or math content and methods classes.
6. Employment of a "clinical model" for science and math methods classes whereby school district faculty and others are involved in the presentation of science or mathematics teaching methodology.
7. Implementation of a comprehensive content/pedagogy-based summer institute series for pre-service science and math teachers and in-service science and math teachers
8. An academic year special support program for final student teachers, teaching interns, and new teachers.
9. Implementation of a program that results in hiring assurances from local district for FCEPT graduates.
10. Implementation of a partnership program that serves as a dissemination model for the nation and one that is fully institutionalized beyond the funding period.

Teacher Profiles

In 2001 there were seventy-five students heavily participating in FCEPT with forty-seven (63%) being minority. This minority participation closely reflects the regional demographics and is evidence for FCEPT meeting its target for underrepresented groups. We believe this data makes us one of the highest minority enrolled single subject science and math teacher preparation programs in the state and nation. The breakdown for the different FCEPT participants shows that fifty-five students are enrolled in the FCEPT program as undergraduates with 56% (31/55) of the students from minority groups and 55% being female (30/55). The Credential Program has twelve students, 75% of whom are minority. Lastly, all five of the first-year teachers are minorities and two of the three interns are minority. The ethnic breakdown of undergraduate FCEPT minority students is 74% Hispanic, 10% Hmong, 6% African-American, and 10% other minority. FCEPT students have a relatively high GPA of 3.11 (S.D.= 0.40) as a group with a median of 3.095. Eighty-nine percent (47/53) of the FCEPT students were enrolled at CSUF and 11% (6/53) were enrolled at FCC.

FCEPT students represent all undergraduate classes (i.e., 3 freshman, 10 sophomore, 18 juniors, 22 seniors, with 1 getting a second degree). The majority of students recruited have a preference for teaching high school, though approximately 40% of the students have yet to decide which grade(s) they prefer to teach.

FCEPT students who are often referred to as FCEPT 'Fellows' are expected to maintain full-time student status, attend all project meetings and events, and maintain a fairly high academic grade point average. In turn support is provided by the project. This support involves summer program support, an academic year scholarship, enrollment in a paid early field experience "Teaching Assistantship Program" and other income earning programs as appropriate. Naturally, FCEPT, to date, has lost some students due to a variety of factors such as career changes, inability to meet academic requirements even after extensive assistance, or failure to maintain a commitment to the project requirements etc. While painful initially, in the long run, our standards have made the project more attractive to potential future science and mathematics teachers and we are looking at a waiting list in the future and/or the expansion of support for these students. The project is not viewed as an entitlement program, which of course it is not. The causes for the attrition are many. However, the project has different programs in place in an attempt to support all students that participate including, but not limited to tutoring, counseling, work placement and others. We are dedicated to losing as few students as possible especially when financial reasons are cited as the major problem. Within the scope of our ability to provide assistance, it is provided. Those students who are dropped are given every opportunity to re-enter if there is evidence that such consideration is warranted.

Special Academic Year and Summer Programs/Institutes

The FCEPT summer science institutes provided our FCEPT science students and mentor science teachers with an opportunity to work together in a relationship and setting very different from the Teaching Assistant/Mentor Teacher role played during the academic year. The summer science institute also provided opportunities for college and university science faculty to field test and revise certain elements for delivery of content of regularly offered courses. During the summer of 1999, FCEPT, merging resources with the Fresno Unified School District's Urban Systemic Initiative, offered two major institutes. The intensive 10 day-long Secondary Science Institute (with a follow-up field trip held later in the summer) focused on integrating key cross-disciplinary science concepts incorporating energy dynamics and environmental topics as major points of connection. FCEPT has also enjoyed outstanding success in providing some very worthwhile enrichment workshops and institutes in science and math/education topics.

We have provided or are currently providing academic year workshops/summer programs dealing with entomology, environmental science, biotechnology, equity in science and math education, laboratory interface software and hardware, applications of technology in teaching science and mathematics, embedded literacy in science and mathematics, fractals and chaos, problem solving, web design in math teaching, and modeling physics. These workshops and institutes have been well enrolled and well received. These programs provide a means by which the university/college faculty and district science and math teachers can pilot teaching strategies. These are often later infused into regular academic year courses[1] taken by undergraduate and credential students in the pre-service program. Evaluations of these offerings have revealed a very high level of satisfaction on the part of the participants.

During the summer of 2000 and 2001, FCEPT offered three institutes: Modeling Physics (MP), Conceptual Chemistry, and Problem Solving or Modeling Mathematics. The institutes were three weeks long and occurred in June of 2000 and 2001. Each institute met Monday through Friday for three hours (i.e., 1:00 P.M. – 4:00 P.M.) and were co-taught between university faculty and area high school teachers. We will focus on the results of the summer physics modeling workshops and its impact on the FCEPT students. In this paper, the term "modeling" reflects the spirit of Arizona State University's Modeling Method of instruction.[2] The following summary from the workshop's evaluator includes candid responses from the participants.

Modeling Physics - Institute Description

Modeling Physics Institute (MPI) is a learning approach that attempts to help students construct conceptual understanding by converting physical models into mental pictures. Instead of the traditional coverage of seemingly disjointed topics, a few core models are used to describe a broad range of physical phenomena. The process involves the development of commonly understood operational definitions, a paradigm laboratory activity, a consensus model, application and refinement, and ultimately deployment. Furthermore the consensus description is expected to be expressed through multiple representations (graphical, verbal, diagrammatic, and mathematical) to reflect a range of learning styles. In each institute the participants have split into two sections based on self-selection using their comfort level with physics as the benchmark. The students more comfortable in physics dealt with force/motion models (FMG); whereas, those uncomfortable with physics worked on graphing and kinematics (GKG). Generally, the high school physics teachers and approximately half of the FCEPT students were in FMG and middle school teachers and the other FCEPT students were in GKG.

Goals

There were both content and pedagogical goals for the modeling physics institute (MPI) with different content goals based on the session's participants. The session instructors explained:

> FMG (Comfortable Group) *...the major [content] goal is to allow the participants to have a better understanding of motion and force as Newton would see it and measure it by the Force Concepts Inventory (FCI), so we pre-tested and post-tested.*

GKG (Uncomfortable Group) ...*they need to have better understanding of motion and I would really like them to have a good understanding of how to read a graph, how to develop a linear model from data. (These students were pre and post-tested using the Test of Understanding Graphs in Kinematics – TUG-K).*

For both groups: *The second major goal is a pedagogical goal. We want them to learn how to employ student centered activities that have the right amount of structure so that students are learning, that they are constructing their own learning, but while at the same time you are getting through content at a reasonable pace. Basically, we try to model, as facilitators, we try to model those teaching techniques that we want them to use.*

Course Design

The MPI blended whole class demonstration/lecture in a lecture hall with experiments carried out in physics student labs. As mentioned previously, the students were separated into two groups based on the students' comfort with physics. The experimental sessions were in small groups doing:

- Data collection based on a guiding set of questions or models that needed explanation.
- Data collection and analysis using computer-based probes and software.
- White boarding (i.e., small group processing on dry erase marker whiteboards).
- Group presentation of results and conclusions.
- Class questions and discussion.
- Deployment questions (i.e., questions that dealt specifically with the lab content.

Lab sessions were facilitated by the instructors that modeled the following:

- Questioning.
- Building knowledge from student understanding.
- Using evidence for conclusions.
- Explaining based on words and mathematical statements.
- Small group work.
- Whole group discussion.

The course was designed to promote a constructivist classroom where the instructor facilitated learning. In an attempt to determine what physics knowledge students had at the beginning of the course, the instructors had the students take two standardized multiple-choice assessments (i.e., Force Concept Inventory (FCI)[3] and Test of Understanding Graphs in Kinematics (TUG-K)[4]). The instructors also had the students re-take the assessments at the end of the course; however, the students that were not comfortable with physics (GKG section) did not re-take the FCI because their curriculum did not focus on force. The results of these assessments will be given below in the student achievement section. The FMG students also took the Force and Motion Concept Evaluation (FMCE)[5] during the course as a second check on student understanding of force and motion.

Course Activities

Each day a 30-45 minute joint session was held with all participants (math, chemistry and physics) introducing a common theme that was part PowerPoint presentation, demonstration and whole group discussion. The themes were chosen from "Hestenes Lectures"[6] ("Expertise", "Preconceptions", "Improving student discourse", "Cognitive foundations", and "What to teach") and from field-expertise provided by the experienced workshop co-leaders ("The modeling cycle", "What is a model?", "Whiteboarding", "Modeling assessment", "Socratic dialog", "Ratio reasoning", and "Dealing with standards"). Students then went to their respective lab rooms. Both the FMG and GKG sessions lasted about 2.5 hours with a break and were populated by 15 and 14 participants respectively.

A typical day involved examination of a paradigm lab activity, through data taking, whiteboard discussion, and workbook activities. At all times the facilitator modeled how an MP teacher would act in the classroom. Below is an example taken from one of these activities emphasizing the point of ratio reasoning. One group of three students presented their findings on an experiment to determine the relationship between constant acceleration and the velocity of an object (car rolling down an inclined ramp). The students described their experimental arrangement of photogates and the data. Their conclusion was "...velocity increases at a steady rate." The facilitator modeled the philosophy and pedagogy of MP by having students define new terms. The following excerpt from this discussion follows:

> Facilitator – *You are using a new term, acceleration, what does this mean?*
> Student – *Change in velocity.*
> Facilitator – *Can you interpret acceleration as a ratio?*
> Student – *For every one second, speed increases by 0.809 m/s.*

After the second group presented, the facilitator stated: "Let's take a couple of minutes to talk about the modeling cycle." The facilitator then used 12 minutes to go over position versus time graphs and velocity versus time graphs for constant velocity particle model. There was discussion about how a teacher would do this in a classroom where friction may be an issue. There was also a discussion about uniformly accelerated particles with associated graphs. In summary, the facilitator restated the modeling cycle:

- Pre-lab discussion– development of operational definitions.
- Collect data on computer laboratory interfaces and analyze with graphing software.
- White boarding discussion.
- Consensus and post-lab discussion.
- Deployment worksheets.

Similar activities took place in the FMG section, with the emphasis more on force models. Prior to, during and after the presentations the facilitator discussed issues about classroom implementation of modeling instruction. For instance, prior to group presentations the facilitator stated:

"When I teach, the first several groups take more time... Think what you agree and disagree with and talk about that."

During the first presentation: "One thing I do [in my classroom] is make a few comments [about the presentation]. After the presentations: "We are going to do consensus, I am pretending to be the teacher… The conversations you are having are the same as kids. I have about 80% of my kids involved, that is great and much better than when I lectured."

When groups presented their results the facilitator also focused on making meaning of the content and graphs. For example, he asked groups: 1) Put your mathematical statement into words; 2) What is the significance of the line going through zero? 3) If you double the mass, what happens to the length? Lastly, closure of the activity centered around the slope of the lines that the different groups generated; what those slopes mean; and if the group found a constant. The consensus was that the slope of the line indicated the area density of a given material and because the groups used different material that there was no general constant. The class ended with a homework assignment. Both sections of the MPI attempted to implement the content using the educational tenets developed by Hestenes et al. The survey that follows will give the student perspective on the frequency of instructional strategies that the facilitators used during the course.

Student Surveys

Students in the MPI filled out the same survey for both summers. The survey attempted to determine the frequency of instructional practices that are considered to be reform oriented. The scale used on the survey is a 4-point Likert with 1 being never and 4 being almost every class. Analysis of the mean responses from 29 students this year indicates that MP sessions used reform oriented instructional strategies on a daily basis. That is, a majority of survey items had a mean rating close to (4). Some of the more important items that the MPI used almost daily are:

- Structured cooperative groups.
- Whole-class discussions.
- Use of objects or models.
- Perform investigations.

A comparison between the two year's responses did not give any significant differences; however, there was a trend for items to be rated slightly lower in 2001. In the table below, students rated 9/12 items lower this year than last with "Write lengthy descriptions of your reasoning" decreasing the most (i.e., 3.2 to 2.8).

When asked about the different frequency of assessments and assignments in the courses, both groups felt there were multiple types of problems, assessments, and assessment items that asked for different information. It is significant, and aligned with reform practices, that this course used a variety of methods to have students engage with the content. For example, MP provided opportunities for students to engage in items involving knowledge and comprehension and items involving application, synthesis, and evaluation almost on a daily basis (rating =3.7 on both items). Again, the ratings for 2001's MPI were less in 9/10 items with the largest decrease observed in writing (i.e., essays 2.6 to 1.8). The MPI used technology almost daily to learn information and to gather and analyze data.

1=Never, 2=Rarely (1-2 times/course), 3=Sometimes (3-4 times/course), 4 = Almost every class

	2001 MPI (n=29)		2000 MPI (n=34)	
	Mean	SEM	Mean	SEM
Work in structured cooperative groups?	4.0	0.03	3.9	0.07
Participate in whole-class discussions?	3.7	0.10	3.9	0.04
Ask content-oriented questions in class?	3.8	0.07	3.9	0.04
Manipulate concrete objects such as models or 3-D representations?	3.7	0.08	3.9	0.06
Do activities where you could pose questions and seek evidence?	3.8	0.08	3.9	0.07
Write lengthy descriptions of your reasoning?	**2.8**	**0.16**	**3.2**	**0.16**
Work on problems related to real world or practical problems	3.3	0.13	3.5	0.15
Perform investigative activities such as data collection, analysis, and various types of representation?	3.9	0.06	3.9	0.05
Make connections to other SMET and non-SMET fields	3.0	0.16	3.3	0.15
Participate in required field experiences that enhanced your knowledge of course concepts?	2.6	0.25	2.4	0.25
Design presentations that facilitate your learning course concepts?	3.6	0.14	3.9	0.06
Evaluate the extent of your own learning?	3.4	0.16	3.6	0.11

Table 1. *4-point Likert survey comparison between summer 2000 and summer 2001 participant responses to instruction in relevant teacher skills. Note the high average responses imply that the workshop utilized these skills on a daily basis.*

Lastly, it is significant that the MPI promoted a supportive culture for reflection and learning. The scores in the following table are indicative of a course that is very much reform oriented and is attempting to become student and inquiry centered and make the student responsible for their learning.

1=Never, 2=Rarely (1-2 times/course), 3=Sometimes (3-4 times/course), 4 = Almost every class

How often did you complete assessments /assignments that include:	2001 MPI (n=29)		2000 MPI (n=34)	
	Mean	SEM	Mean	SEM
Problems with straightforward (simple) solutions?	3.4	0.14	3.4	0.12
Problems with complex rather than simple solutions	3.4	0.13	3.6	0.10
Portfolios	1.7	0.23	1.8	0.19
Essays	**1.8**	**0.17**	**2.6**	**0.19**
Multiple choice-Short Answer items	2.2	0.18	2.3	0.18
Items involving knowledge and comprehension	3.7	0.10	3.9	0.07
Items involving application, synthesis, and evaluation	3.7	0.10	3.9	0.06
In-class activities that helped you gauge how much you understood course concepts	3.7	0.11	3.9	0.06
Assessments that required you to think about concepts and critically analyze information	3.6	0.12	3.7	0.11
Assignments that required you to think about concepts and critically analyze information	3.7	0.10	3.9	0.05

Table 2. *4-point Likert survey comparison between summer 2000 and summer 2001 participant responses to assessment issues, affirming the workshop was geared toward active engagement.*

Although the MPIs provided opportunities aligned with the reform course indicators, comparison between the two MPIs exhibited a decrease in 10/11 items. Three of the reductions were 0.4 or greater and were in indicators that the reform most wants to increase (i.e., questions in bold). The survey data provide evidence supporting classroom observations indicating MP uses:

1=Never, 2=Rarely (1-2 times/course), 3=Sometimes (3-4 times/course), 4 = Almost every class				
Use Technology (e.g., computers)	2001 MPI (n=29)		2000 MPI (n=34)	
	Mean	SEM	Mean	SEM
To learn more information	3.5	0.13	3.5	0.13
To drill and practice skills learned in class	**2.8**	**0.22**	**3.2**	**0.18**
To understand in more depth concerns taught in class?	3.6	0.15	3.8	0.08
As a tool in investigations to gather and analyze scientific or mathematical data?	3.7	0.10	3.9	0.05
As a tool to prepare written reports or presentations	2.8	0.21	2.7	0.22

Table 3. *4-point Likert survey comparison between summer 2and summer 2001 participant responses to technology usage, which were generally supportive.*

1=Never, 2=Rarely (1-2 times/course), 3=Sometimes (3-4 times/course), 4 = Almost every class				
Course Culture and Philosophy of Learning	2001 MPI (n=29)		2000 MPI (n=34)	
	Mean	SEM	Mean	SEM
Have opportunities for reflection in mathematics/science?	**3.3**	**0.17**	**3.7**	**0.10**
Supportive atmosphere for learning new ideas?	**3.5**	**0.18**	**3.9**	**0.07**
Learning settings that showed respect for diversity	3.3	0.16	3.6	0.13
Learning settings where you influenced course activities	**3.1**	**0.16**	**3.6**	**0.11**
Asked you to take responsibility for your own learning	3.6	0.13	3.8	0.08
New information was based on what you already knew about the topic	3.4	0.16	3.5	0.12
You learned course concepts through inquiry	3.6	0.11	3.9	0.05
The instructor acted as a guide to facilitate learning	3.8	0.09	3.9	0.07
Your assessment results are used to modify what is taught and how	3.3	0.17	3.3	0.16
You have appropriate time, space, and resources to learn the course concepts	3.5	0.17	3.8	0.09
Classes are communities of learners	3.6	0.16	3.8	0.12

Table 4. *4-point Likert survey comparison between summer 2000 and summer 2001 participant responses to culture and philosophy, supporting the reform-oriented workshop.*

- Reform-oriented instructional strategies that are implemented on a daily basis.
- Multiple methods to assess student understanding that are implemented during the course.
- Technology to collect and analyze data and as a tool to help learning.
- Inquiry and student-centered practices.

The data also indicates that there may be less of many of the above occurring in any given day, especially in asking students to write. This may be something that the instructors monitor in subsequent institutes. It is one thing to change instructional practice and another to increase student understanding. The MP students took several assessments (FCI and TUG-K) at the start and end of the course. These results are presented next.

Student Achievement

The MP students were given the Force Concept Inventory[1] (a 30 item multiple-choice test developed to assess student understanding of Newtonian dynamics and the Test of Understanding Graphs in Kinematics[2] (a 20 item multiple-choice test developed for testing the ability to interpret graphical representations of motion). Participants were pre-tested and post-tested on the FCI and TUG-K with the exception of the GKG students not taking the FCI post-test. The GKG students did not

TUG-K	All Students (n=30)	FMG (n=15)	GKG (n=15)
Pre-Test (mean ± SEM)	9.0 ± 0.99	12.1 ± 1.30	6.0 ± 1.03
Post-test (mean ± SEM)	13.1 ± 0.79	14.8 ± 1.03	11.4 ± 1.04
Growth (Raw Score ± SEM)	4.1 ± 0.53*	2.7 ± 0.71*	5.4 ± 0.67*
Growth (Percent change x 100)	46	22	90
Growth (Normalized x 100)	37	34	38

Table 5. *TUG-K results for all students, FMG students and GKG students. (*) Denotes significant difference at $p<.05$, two-tailed paired-sample T-test. Percent growth is calculated by dividing the increase in mean raw score by the pre-test mean. Normalized growth is calculated by dividing the mean raw score by the difference between the maximum score minus the mean pre-test score (e.g. $4.1 \div (20 - 9) \times 100 = 37$).*

study forces during the course; thus, there would not be any reason to see great improvement on the FCI. The results of the TUG-K indicate significant growth in students' ability to interpret graphical information about motion over the 3-week course. There are three important points to make about the results:

1. There was significant average growth based on all categories. Not shown is individual data that indicates every student increased their score with one exception – a student that pre-tested at 19/20 post-tested at 18/20.
2. FMG and GKG groups represent students with different experience and knowledge with respect to interpreting graphs dealing with motion. That is, pre-test data comparison shows that the average FMG student (12.1) had twice as many correct answers as did the GKG group (6.0).
3. GKG students increased their scores the most, i.e., from 6 to 11.4 correct answers but there still is a significant difference between the FMG (14.8) and GKG (11.4) student scores as a group.

Another way to look at the data is to break the groups into FCEPT students, teachers, middle school teachers and high school teachers to see if there are any differences between these groups.

The following table provides data on FCEPT students, all teachers and teachers separated into middle and high school. Although no significant differences were found, there is a trend for FCEPT students and high school teachers to score better on the TUG-K than middle school teachers. The most important observation to make from this data is that: All groups showed increases in their scores, indicating that the course is meeting some of the needs of all participants. Changes between pre- and post-test scores were not significantly different due to the small sample size and relatively large variation in scores within a subgroup.

The TUG-K data appears to demonstrate that some of the goals of the MPI were met with respect to enhancing student knowledge and understanding of interpreting graphical data pertaining to motion, similar data was found for the FMG students on the FCI.

TUG-K	FCEPT Students (n=18)	Teachers (n=12)	Middle School (n=7)	High School (n=5)
Pre-Test (mean ± SEM)	10.3 ± 1.32	7.2 ± 1.38	6.1 ± 1.75	8.8 ± 2.18
Post-test (mean ± SEM)	13.7± 1.05	12.2 ± 1.17	10.6 ± 1.41	14.6 ± 1.44
Growth (Raw Score ± SEM)	3.4 ± 0.66	4.6 ± 0.90	4.4 ± 1.13	5.8 ± 1.46
Growth (Percent change x 100)	33	64	72	66
Growth (Normalized x 100)	35	36	31	52

Table 6. *Pre and post TUG-K data for FCEPT students and teachers. No significant differences were observed between pre- and post-test scores or between groups. Percent growth is calculated by dividing the increase in mean raw score by the pre-test mean. Normalized growth is calculated by dividing the mean raw score by the difference between the maximum score minus the mean pre-test score.*

As stated earlier, the FCI attempts to measure an individual's understanding of Newtonian dynamics. As a group, the FMG students pre-tested at the 52% level (i.e., pre-test/maximum score; 15.7/30). **The class averaged an increase of 7 correct answers on the post-test with the mean score being 75% — or a 23% overall increase.** FCEPT students initially out-performed the teachers on the pre-test; however, the teachers had higher post-test gains that resulted in both groups reaching the same 75% score on the FCI.

FCI FGM students only	ALL FGM	FCEPT-FGM	Teachers-FGM
Pre-Test (mean ± SEM)	15.7 ± 1.97	17.3 ± 2.66	13.7 ± 2.96
Post-test (mean ± SEM)	22.6 ± 1.25	22.6 ± 1.89	22.7 ± 1.69
Growth (Raw Score ± SEM)	6.9 ± 1.14 *	5.4 ± 1.43*	9.0 ± 1.61*
Growth (Percent change x 100)	44	31	66
Growth (Normalized x 100)	48	43	55

Table 7. *Force Concept Inventory scores pre- and post-test for all FMG students and for FCEPT students and teachers. Percent growth is calculated by dividing the increase in mean raw score by the pre-test mean. Normalized growth is calculated by dividing the mean raw score by the difference between the maximum score minus the mean pre-test score. (*) Denotes significant at the $p<0.05$ level on two-tailed paired-sample t-test.*

The TUG-K and FCI assessment results both increased significantly demonstrating marked students' understanding based on these instruments. Although care should always be considered when using single tests to evaluate changes in student achievement gains, the fact that both assessment instruments showed student gains increases the probability that the instruction during the course contributed to student increase in knowledge. It is important to mention that a part of the success of the MPI in increasing student achievement, based on the FCI and TUG-K, is due to the alignment of curriculum with these assessments. One measure of student achievement from the institute was how classroom practice was impacted. Some of the indicators that will be used were defined by one of the MP instructors:

> *If you walked into a room you would see students engaged in activities, students would be looking at phenomena, asking questions, going through the scientific process and trying to answer those questions. They would be taking data, they would be interpreting the data, developing graphical and mathematical models from that data. They would be presenting their results to the class on white boards. Students would be asking students about what is on their white board. The teacher would not be the dominant source of knowledge, the students would be constructing knowledge, the teacher that is there would be there to hopefully prevent them from heading down deadends, as a guide, as a facilitator. But they would be asking students questions, they would be involving dialogue about the material.*

The next section will highlight a student's experience with both the MP and traditional physics course.

One Student's Story

One of the easiest ways to portray the difference between traditional and reform courses is to find an individual that has experienced both and have them tell you their story of the two courses. This section will provide such a story. A student in the GKG section of MP had just finished first semester physics (2A) at CSU Fresno. We will present her story of the differences she found between the MP and Physics 2A course.

> *I took Physics 2A (algebra-based college physics) this past semester so I just finished and I am taking the Modeling Physics course now. And in Physics 2A, we went so fast, we went over so many terms and I learned how to solve the problem, but I didn't really understand the meaning... I got an A, just because a lot of the physics was math and I was able to solve the problems. [In] the Modeling Physics ...I understand what the terms mean, what position means, using the operational definition, what I think a term means and then using a model to understand what the terms mean and how they are related in graphs... it has brought a whole understanding to everything that I have learned to this point in my physics class that I didn't understand before.*

We start with this comment because it is a powerful statement about how reform instructional practice can impact student understanding of content. This comment is from a FCEPT student that took CSU Fresno's traditional undergraduate physics course and earned an (A) due to mathematical skills – not because of understanding the physics. In fact, this student had a pre-TUG-K score of 9/20 and decided to be in the physics uncomfortable group (GKG) even after obtaining an (A) in introductory physics. This student's post-TUG-K score was 18/20; thus, there is empirical evidence to support the student's statements about learning in the MPI. This student continues with other differences between the MP and traditional physics labs and course:

First of all, in the Physics 2A course, my teacher did do some good experiments in front of the class, but we weren't really involved in the experiments. We were watching, so for me it didn't do very much for my understanding. In the labs [Physics 2A], it was hands-on, but I think, we had the whole experiment written out in front of us. We had to follow the experiment, get the results. ... I may have understood some stuff and I may not have, but in the Modeling Physics, with the V diagram[1], we are creating our experiment, we are solving it, we are getting our own data and that is really helping me understand what is happening and why it happens. So I think it is more, in the Physics 2A lab, it was hands-on, but I wasn't working to understand what I was doing.

The student was asked to expand on some of the strengths of the MPI from her perspective.

One thing that has really helped I think, I really like the modeling. Well first we are doing, in the Modeling Physics, we are doing a V diagram, so we are able to look at the problem statement, find the variables and just write out our little experiment. We are writing out what we think is our hypothesis, what we think the graph may look like, our controls, our variables and the data. Then we are going in and finding the data. We are using computers, but what I really like is the use of, after doing the V diagram, the use in groups of the white boards, because we are able to talk and we are able to solve the problems. What really helps me is to speak about the problems. If I can speak about it and voice it and talk to others and we are working together to solve something and figure something out. I learn it a lot better than if a teacher were to just tell me, "This is how it is. This is how you solve the problem. Now do it."

The MPI definitely seemed to allow this student to reflect on the content with evidence that she obtained through experimentation. The student discussed an example of how students approach specific content in the Physics 2A and in the MPI.

Okay, for example, the kinematic equation, like position = velocity x time plus the initial velocity. We did that in the Physics 2A course. He gave it to us. I knew if I had velocity, I plug it in here, if I have position I plug it in here and if I have time and if I have everything but one variable, I take that variable and I plug it in. So I understood that.

But now in this class [MP], because we started with operational definitions and we were able to define things, our own way and actually, what I liked, he had everybody in the class write down an operational definition and then we put them all up on the board, and we saw that they were all different. So because they were all different, they were all kind of in a similar direction, but you know operational definitions are kind of what you think something should be. It helped me learn more because I wasn't following someone else's definition, ...I was learning by what I was doing and what I was thinking, I was working my mind, I wasn't doing what someone else said.

But anyways, for that equation, so we defined stuff and then, what we did is, we started out with graphical analysis... we got the data points, we graphed them and from that, we were learning position versus time. Then we went into learning, the slope of that is velocity and why it is velocity... We related it to the mathematical model, $y = mx + b$ and that is how we got the position, the velocity, the time and the initial position. We kind of related it to that and related it to the models that we were doing. I don't know, just all of that work in relating things helped me to understand where everything was coming from and how to use it in different situations.

One of the reform's most important goal is to have students generalize their findings and understanding so that when different problems surface the student is not looking for formulas in which to plug in the variables, but, instead are looking to understand the general principals that may apply to a problem and work from there. The above statement certainly seems to support the idea that the MPI helped this student move from trying to plug variables into formulas to understanding general principals. Another goal of the reform is to have students work with each other to solve problems and present their results.

We have our partner, but then we interact with other groups. I work with my partner, we talk about things and then we see what the other groups did and compare problems. ...I have talked to some people in the class and I think they feel basically about the same way that I do, it is really helping them in their understanding, maybe not in physics but even in math, concepts. The modeling has helped them understand why we did things in math. Why we are doing things in physics.

MPI had two goals: content and modeling best practice. The following response from the student was received when she was asked if the MPI would make her a better teacher.

I am in the FCEPT program and this last semester was my first semester and I had a slow start getting into the high school. I was in the high school for one month and that was it. So I have had no, or hardly any experience in the classroom and I wonder if I will be a good teacher. But just this course and watching my teacher [MP] teach and the methods he is using, ...It is really helping me get confidence, I can do this, this is really working for me, why can't it work for other students?

The student believes that teaching the MP approach will take more time.

I think it would take a lot more work, a lot more work planning the strategies, planning the modeling or *the experiments, that they want the classes to do. Yes, setting everything up will take a lot of time, I am sure.*

Lastly, the student talks about what she likes most about the MPI versus the Physics 2A.

> *...in my Physics 2A class, it was a bigger class, but the teacher wouldn't really challenge us, he just lectured, he wouldn't ask questions, he wouldn't challenge what we know. And in this class [MP], the teacher is continually asking us questions, making us think and he is not telling us the answer, but he is making us think about it through this process. We are trying to figure it out, and I really liked that, it makes me think.*

This student's story is exactly that – a single person's perspective of the differences between a traditional undergraduate physics course and the MPI. This student believes she learned a lot about mathematics, physics and how to teach from the course. Some of the other students interviewed have a different perception of the two MP sessions. Their comments will follow.

Participant Comments

The observation of three sessions and the student's perception above portray the MPI as being very successful at engaging participants in physics or mathematics in a reform-oriented fashion. The scores from the TUG-K and FCI also indicate that the MPI helped participants learn graphical analysis and physics, respectively. The following comments were obtained from a randomly selected group of participants – three were in the FMG and one was in the GKG. The individual pre-, post-test scores for the 3 FMG participants on the TUG-K and FCI are: (TUG-K: 10,12; 15,16; 28,27; FCI: 7,17; 21,28; 28,27). The GKG participant had pre-post TUG-K scores of 5 and 14. Two of the FMG participants scores increased dramatically on the FCI (one participant had no where to go on this assessment) and the GKG participant exhibited a large increase on the TUG-K. Generally, these participants appeared to increase their knowledge when assessed by the instruments used in the MPI. The remainder of this section will provide quotes from participants about:

- General feelings toward the course.
- Things that they liked about the course.
- Perceptions of impact on their content knowledge.
- Perceptions of impact on their teaching.
- Concerns of Participants.
- Ways to improve the course.

General Feelings Toward the Course

There was a wide range between how individuals felt about the course. Comments from the FGM course first followed by the GKG comments.

> FGM: *Comparing this year and last year, there have been some bright moments, some good moments in the modeling technique and there have been some real downer moments. I understand the whole idea behind Socratic questioning, ...I understand the value of students discovering, if you discover something on your own you really do get it forever. ...but there is also a large degree of frustration when you don't get it. I think one of the flaws in the modeling method is that it is just a little too rigid because when a student doesn't get it, sometimes they just need to be told and then go back to the modeling method and maybe they will get it. But when you are constantly clinging to your Socratic method, sometimes it causes more confusion and more frustration than I think is good.*
>
> *Generally speaking, I feel there is a lot of value in the course, but let me clarify some things. ...what I see is the Socratic method, the modeling method. Those are very good techniques, but there are also techniques and strategies that have to go along with that process, some basic universal techniques that aren't also being modeled and I use that in a different sense, as the Modeling Physics, as we use that word model. For instance, you need to make sure that the techniques that you incorporate when you are giving information to students that they can read that information, that it is legible on an overhead projector, ... If you have equipment that you are using in the classroom, that that teacher has already used the equipment. So many times we have tried to use equipment in a classroom and our instructor has never used the apparatus. ...it has been very frustrating when I spent two hours on a piece of machinery trying to jury rig it to work right and finally coming up with something that might work okay. ...The other thing is, the misconceptions. When you don't have the vocabulary, coming from the first physics class last year, that does not prepare you at all to come into the second physics class. ...I think there are some big holes that never, ever get filled from the first class to the second class.*
>
> *I have a generally favorable impression of the class. It is hard for me to make any kind of judgement, because I don't have any teaching experience, I only have one semester as a teaching assistant and I have never taken any classes on teaching itself, ...I generally like what we are being taught so far.*
>
> GKG: *It is not a course that is going to straighten out the content of physics for you. ...I went in what they said was the less intimidating section, but we have several college kids in there that have just finished physics, just finished the first semester of physics and they struggled with some things too, because it doesn't cover content. ...a lot of the teachers in there have not had any physics and you can tell. They are using velocity and all of these different things interchangeably and you start talking to them and they have really nutty thinking and it is not clearing up.*

The above comments suggest that some participants felt the MP sections were unevenly implemented for several reasons:

1. Implementation of the modeling instructional strategies.
2. Frustration with understanding the content.
3. Frustration with setting-up and using the equipment.

4. Lack of physics content.

Though participants stated frustrations with the MPI most found portions that they liked.

Things That They Liked About the Institute

Most participants enjoyed some aspect of the MPI, though one participant out of 29 could not state something that they enjoyed.

> *I had a student that was with me last year in the FCEPT program, and he happens to be also in the class. I think it is a really nice chance for him to see his teacher in a different light, which is really good, and to also be able to see us side by side as professionals.*
>
> *I like all of the ideas we have done for how to represent physics or how to model physics to students, at this point it is all new to me and it is overwhelming sometimes, all of the ideas. I went home and I taught about the concept of normal force to my ten-year-old cousin and so I wanted to see if I could get the idea across to him. It worked and I liked mostly ideas about how to get stuff across to students, the different ways.*
>
> *I loved being in the class with a lot of different teachers and some students too, that is fun. I love the graphing, but I am very visual. I taught math and I am going to be a more effective teacher. I have always taught graphing and done a lot with it, but I will be a lot better next year…*

Participants enjoyed:

- The interaction between FCEPT students and teachers.
- Modeling approach to physics.
- Modeling approach to graphical analysis.

Perceptions of Impact on Their Content Knowledge

One of the main goals of the MPI is to impact participant knowledge on graphical analysis and physics. The quotes above did not indicate that participants perceived that they learned much from being involved in the MPI; however, the assessment data presented above on the participants indicates growth in the content areas on which each session focused. The participants respond directly to a question asking about growth in content knowledge.

> *I really do understand what is underneath the graph a lot better… I understand what that area beneath the graph means, although sometimes I use the word distance instead of displacement, other than that I am real clear.*
>
> *I always kind of knew the 3 laws of motion, but there are problems that we had to do, or made to think about, mainly in situations that forced us to have a better understanding. For instance the elevator problem. What is the net force on this person riding the elevator or a rocket, or bicycle and what about friction? I guess I just have a better understanding of those 3 laws.*
>
> *I really feel I have a firm understanding of the difference between a change in position, what a constant velocity is and what a changing velocity is and what acceleration is. We had all of that last summer and we are re-going over it again, so I really feel I have a firm grasp of that. If I went back to the middle school and I had to teach just that basic element of motion, velocity, acceleration, I feel like I could do a really good job. …in terms of Newtonian physics and forces, how that all plays in physics, I think I am finding out that what I thought I understood I really don't understand. I feel more confused with force, which maybe is a good thing, because maybe philosophically, when you really start learning something, first you have to figure out what you don't know and I have realized that I just don't understand force. I don't have a firm concept, especially when we have to do the force diagrams: What are the relevant forces? What are not? How they all sum together to create some type of a system. I feel like at least I am aware of my non-understandings, so I could say I have benefited by coming aware of my ignorance on the subject.*
>
> *...the beginning motion models I feel comfortable with. Then we started working on our force models and I feel like I have learned a lot to be honest with you, even though it is learning through frustration. I still feel that through scratching and clawing, I have come up with learning some of the mathematical concepts of the force models which is good. Even today, I got more clarification for that and so that is making me feel a little bit better. I feel that I can relate some of the forces better on my own personal experiences than I would ever have last year, in fact I didn't really understand how the sum of forces, taking a look at all of the forces that affect a system, how that affects the physics of things. I am getting some concepts there. …But concepts are being taught sometimes very quickly.*

Regardless of some of the concerns expressed by the participants, all believed that the MPI impacted their content knowledge. T hey also believe that it has or will help them teach motion, velocity and acceleration to their students.

Perceptions of Impact on Teaching

The second major goal of the MPI is to provide pre-service students and teachers with instructional strategies that will allow them to teach for understanding. The following responses suggest that the participants' teaching has already or may be impacted by the course.

> *Now that I will say, because I am going to use that [white boarding]. I like the white boarding and I like the fact that kids have to say out loud what they are thinking, that is real good and they love to perform. The old common name for modeling was constructivism… The whole thing about given the experiment that has real narrow parameters, let the kids do it within a certain time, let them explain it and then let's have some consensus about the law… It is great teaching, it always has been. And I will use it and I will use the equipment.*

Definitely being at it for 2 years, I can say that after last summer's course, my students did white boarding almost the whole year. My kids like white boarding, it is a great way to find out what students are actually learning. I think it is a great tool for assessment. I think it also develops their speaking skills, they start out pretty shaky, but you can see the improvement. ...I think the white boarding helps good critical thinking. I think this class has taught me how important it is to prepare, which I already did know, which any science teacher knows, if you are not prepared, if you don't have the equipment ready and things just fall apart. This summer in particular though, what I have learned the most is what it feels like to be a student who is in the lower 10 or 20% of the class and I haven't felt that in ages. I think the last time I felt that was in a 7th grade math class when I was a kid. I understand now the value of the teacher needing to be aware when students are not grasping what he or she is presenting and that teacher needs to address those needs. I don't feel they have been addressed here and it makes me realize, you know, I need to be more sensitive as a teacher myself. I need to make sure that all of my kids are understanding and if they are not, I need to figure out a way to get them to understand.

...I am definitely using the white boards, I will continue to use the white boards in my classroom. I think they are invaluable. It really does help the students to be able to communicate and to feel, the way I use them in my classroom, they feel safe, it is a safe environment. I also use bits and pieces of Socratic questioning. ...Also just the process of hands-on activities. I have always been a hands-on oriented teacher, but it just validates it even for higher level.

I think the white boarding is the most valuable thing that will probably help me when I start teaching. ...and in fact the students have to explain themselves and the fact that it is important to have a safe environment, or it is okay to be wrong.

The participants believe that the MPI has and will impact their teaching in the following ways:

- White boarding as a tool for students to think critically.
- Group discussion and presentations as a method for students to improve communication skills and understanding of content.
- Make teacher more sensitive to struggling students and implement different strategies for these students.
- Use guided experimentation, with limited time and closure on concept.

The MPI appears to have had a positive impact on participants' pedagogical content knowledge.

Concerns of participants: Instead of quoting the participants in this section, summative concerns are discussed to help provide guidance for future institutes. The stated concerns include:

1. Modeling process was not adhered to and major steps in developing terminology, definitions, experimentation and consensus was shortened.
2. FCEPT students did not see the modeling process in its entirety and may not be as prepared as they could be to implement modeling in classrooms.
3. Safe environment for all participants to engage in discussions was not created, high-end students sometimes dominated discussions and questioning time.
4. Consensus was not always done and should be major part to clear up confusion.
5. Worksheets and directions to complete them were not always clear.
6. Primary instruction was interrupted by co-instructor in a manner that did not address issue in a modeling format – discussions were occasionally between co-instructor and one or two students.
7. Lecture and demonstration portion of course seemed antithetical to modeling approach and some topics did not seem to connect to course (e.g., analogy versus metaphor).
8. Instructors were not always prepared to use and teach participants on the experimental equipment.
9. Socratic method may have limitations with equipment use and problems when working with time constraints of course.

After the participants expressed their concerns about the course, they also provided their thoughts on what they believe will improve the course.

Participant Suggestions on Ways to Improve the Course

- Have instructors better prepared on equipment set-up and use.
- Provide historical perspectives and stories of discovery of motion and force.
- Start each lesson with 30-40 minute demonstration to uncover or dispel misconceptions (e.g., provide some examples of misconceptions and show some conceptual physics examples) and link to experiments participants are doing in lab.
- Provide answers or textbooks that have answers to some of the questions that are raised – use more than modeling.
- Do not have participants return for more MP – one is enough.

The participants that were interviewed had very different backgrounds and knowledge of physics. Many of the concerns and suggestions came from middle school teachers and the MP instructors need to take that into consideration when they reflect on these issues and the impact on particular course sections (i.e., FMG and GKG). It is important to note that the above suggestions are opinions not always supported by research findings. For example the very last opinion suggests only a single MPI was needed to prepare teachers for the classroom. Extensive research has shown that continued professional development is essential to develop the necessary expertise for successful instructional practice.

Summary

The Modeling Physics course had many positive outcomes this past summer, including:

- Significant student gains on physics concepts and graphical interpretation based on the FCI and TUG-K pre- and post-assessments;
- Modeling instructional practices (e.g., white boarding, small groups, model building and investigative science) though these appeared to decline slightly when compared to last year's MPI;
- Participants' use of modeling strategies in their classrooms (i.e., based on instructor and past participant interviews); and
- Instructors' collaboration in implementing MPI.

The observations of the evaluator support many of participants' beliefs about the MPI. That is, when he observed both sections instructors were using the modeling approach to address content and pedagogical goals, students were doing experiments, white boarding in small groups and presenting results to the class. In most of the sessions he observed the instructors explicitly stated the instructional strategies they were using and how they worked for them in the classroom; hence, they were modeling best practice and exhibiting change in their own classroom practice.

The evaluator also observed instances when the modeling was not followed, when students that understood the content better, dominated sessions and when the co-instructor interrupted the primary instructor without using the modeling philosophy. There definitely was some inconsistency between message and method. With all of these observations, however, there is no denying that almost every single MP participant showed significant gains on either the TUG-K, FCI or both. The student whose story was presented is simultaneously a powerful testament to MP and an indictment of traditional physics.

From the evaluator's perspective, many of the concerns of MP expressed by the teacher participants that were interviewed illustrates the same dichotomy posed by the MP and traditional physics course, that is, several of the teachers stated that all they needed was to know enough physics to teach to their students, they did not necessarily want to take the time to understand the physics concepts – though they apparently understood them better after the MPI.

Overall Impact of Modeling Science on Central Valley Teacher Participants

We have learned that several of our region's veteran science teachers or preservice teachers now teaching under contract, are using the modeling approach rather extensively in their classrooms. The Socratic dialogue strategy combined with the use of white boards to encourage student discussion and to elicit student reflection is widely seen in classrooms. Our teachers have indicated to us that a significantly greater percentage of their students are engaged in the activities and discussions. They attribute this success to their use of the modeling approach. Several teachers have stated that they are able to successfully link the modeling approach to the California Standards. Additionally, some of our teachers have formed their own modeling networks and share their successes and challenges in using modeling in the classroom. This has proven to be a very positive, unanticipated development. Other teachers have stated that, overall, their classroom practice has been greatly influenced by their training in modeling. They report that they are far more student-centered, approach the teaching of science concepts from an inquiry basis, and engage a wider range of students in classroom discussions. Teacher also report that many of their students have shown increased skills in the interpretation of graphs and diagrams as a result of the application of modeling principles in their teaching. Finally, and we think importantly, many teachers are reporting that their students appear to be considerably more enthusiastic about their science classes as a result of modeling. All in all, the program is beginning to impact valley schools and we hope that modeling turns on more students on to science and to considering science majors (and possibly science teaching) as career choices.

Bibliography

[1] J. Vesenka, et al., "A Comparison between traditional and "modeling" approaches to undergraduate physics instruction at two university with implication for improving physics teacher preparation." *J. Phys. Tchr. Educ. Online, 1(1),* 3 (2002).
[2] http://modeling.la.asu.edu
[3] http://modeling.la.asu.edu/R&E/Research.html
[4] www2.ncsu.edu/ncsu/pams/physics/Physics_Ed/Articles/TUGKArticle.pdf
[5] http://physics.dickinson.edu/~wp_web/wp_resources/wp_assessment.html
[6] http://modeling.la.asu.edu/modeling-HS.html
[7] For details of this modeling construct contact: JerryB3118@aol.com

A model for preparing preservice physics teachers using inquiry-based methods

Michael Jabot
School of Education
State University of New York at Fredonia
Fredonia, NY 14063
jabot@fredonia.edu

Christian H. Kautz
Physics Department
Syracuse University
Syracuse, NY 13244
kautz@physics.syr.edu

This paper characterizes the possible impact of using a guided-inquiry approach to the teaching and preparation of physic teachers. The study described in the article compares two groups of high-school students. One group of 74 students was taught using a process of guided inquiry as the main mode of instruction. A second group of 55 students was taught in atraditional, lecture- and laboratory-based format. Written pre- and post-tests were used to assess student understanding of the topic before and after instruction. Results from these tests suggest substantially greater learning gains by the group that had been taught in the guided-inquiry format. The implications of using such an approach to the preparation of future physics teachers in then presented.

Introduction

As we seek to better train future physics teachers, those of us involved in teacher preparation must begin to focus on developing practices that begin to answer some very difficult questions concerning our education system. It is important to realize that we have seen a generation of students graduate from our high schools since we first began to focus on a performance-based system of education. Are our new physics teachers any better educated than their predecessors? If not, where is the breakdown in our system? Do we prepare teacher candidates to be results-oriented knowing that they are likely to be employed in an environment where performance is not the measure of student success? Are we attempting to align new standards with the old practices, rules, and regulations that have traditionally defined education? How do we move our understanding of performance from theory to practice? One thing is certain, if those of us involved in teacher preparation do not choose to answer these types of questions, we can be assured that those involved in deciding public policy will (i.e. witness the policies set forth based on the No Child Left Behind Legislation).

This paper first reports on a study aimed at comparing the outcomes of two different methodological approaches to the teaching of heat and temperature. The study involved two groups of high-school students. One group of 74 students was taught using a process of guided inquiry as the main mode of instruction. A second group of 55 students was taught in a traditional, lecture- and laboratory-based format. Written pre- and post-tests were used to assess student understanding of the topic before and after instruction. Results from these tests suggest substantially greater learning gains by the group that had been taught in the guided-inquiry format. This paper concludes with a discussion of the implications that this type of study has on teacher preparation.

Before moving forward it is important that a description of the two teaching situations be given. What follows is a brief description of the overall approach to teaching in each of the classrooms.

In the "traditional approach", the students were assigned to a 40-minute block of class time five days a week and received a 40-minute laboratory session twice a week. During the scheduled class time a typical teaching episode included the teacher reviewing answers to the previous nights homework; a brief session of student questions concerning the previous days topic; the teacher introducing the next stage in the development of the topic; and finally guided practice of the next set of problems assigned for homework. The laboratory experiences for the students were confirmation labs concerning the topic of specific heat of different materials. At the time of this study the teacher teaching this class had been teaching physics for 20 years.

In the "guided-inquiry" classroom, students were also assigned to a 40-minute block of class time five days a week and received a 40-minute laboratory session twice a week. During the scheduled class time, the students were assigned a task to complete, for example, "at what temperature does water boil?". The students worked in small groups to answer this task. Following this the students were asked to whiteboard their responses and share their results with the other groups. Any discrepancies between the groups data was then used as a point of discussion lead by the teacher. Any homework assignments focused on writing a more formal explanation of what the whole-group consensus was for the discussion. These explanations by students were used to document the students' changes in understanding and student responses were used as the teacher facilitated future discussions. The laboratory sessions of the guided inquiry group followed the same format as the scheduled class sessions. It was noted by students that what they were doing during the laboratory sessions did not seem to be any different that what they were doing during class. The teacher in this classroom had been teaching physics for 12 years.

Overview of study

There is an extensive body of research on student understanding of heat and temperature at the pre-college level.

(See, for example, Erickson,1979, Stavy and Berkovitz, 1980, Nachmias, Stavy and Avrams, 1990. Reviews of various studies are given by Erickson and Tiberghien 1985, and by Driver, Squires, Rushworth and Wood-Robinson, 1994.) Much of this research identified specific difficulties that are prevalent among the populations studied. Other investigations have focused specifically on how student ideas of these topics change as a result of teaching. (See, in particular, Part B of Erickson and Tiberghien, 1985.) A study by Rosenquist revealed that similar student difficulties are often still present at the introductory college level. These results served as a basis for the development of specific curriculum designed to address the identified difficulties (Rosenquist, 1982).

Detailed investigations of students' ideas about a given topic, and the consecutive design of curriculum, however, are generally beyond the realm of an individual secondary-level teacher. The present study, therefore, attempted to find out to what extent the adoption of an inquiry-based teaching style, starting from presently available instructional materials, can help improve high-school teaching of the topics of heat and temperature. A pretest on various topics in thermal physics was administered to two groups of students (N_1=55, N_2=74) at two suburban high schools in Central New York State. By administering a pretest, we sought to determine the level of student understanding of relevant concepts before instruction and thereby ascertain the comparability of the two groups. Some of the pretest items (including the ones on the specific topics of temperature and heat transfer described in Section III below) were based on questions taken from the Heat & Temperature section of Tools for Scientific Thinking (Sokoloff and Thornton, 1997). These multiple-choice questions had been used successfully to elicit common student difficulties identified by physics education research. The format of the pretest used in our study was an 'enhanced' multiple-choice format in which students were required to give explanations for the answers chosen.

For both groups, instruction on heat and temperature began within a few days after the administration of the pretest and concentrated on the topics outlined in the pertaining portions of the Internal Energy unit of the New York State Physics Syllabus (NYSED, 1987). This unit includes the topics of Temperature, Internal Energy and Heat, Kinetic Theory of Gases, and Laws of Thermodynamics.

One of the groups in the study was given traditional lecture-based instruction (N_1=55). The other was given instruction based on a guided-inquiry approach (N_2=74). The instructional activities used with the latter group were derived from Physics By Inquiry (McDermott et al., 1996), a set of laboratory-based modules designed to help students develop a conceptual understanding of the course material and intended primarily for the preparation of pre-college teachers. Modifications were made to the activities to allow for use in a typical high-school setting. A brief description of the activities is included in Section IV below.

For both groups, instruction was completed within a two-week period. A post-test was administered to each group following completion of instruction. In order to rule out improved performance on the post-test due to repetition or rote memorization, we chose to use different questions that required application of the same concepts and ideas. Since a detailed study of the equivalence of two tests in measuring the same aspects of student understanding was beyond the scope of this project, we selected matching items, where possible, that differed only with respect to surface features.

In our analysis of pre- and post-test responses we paid particular attention to the reasoning presented by students in their written explanations.

Test Instruments

On each of the two tests, four matched questions were asked to assess the students' understanding of important ideas related to heat and temperature (H&T). (Each pre- and post-test consisted of a total of eight questions). The remaining questions on the tests that are not discussed here concerned the following topics: the latent heat of melting, Avogadro's hypothesis, the ideal gas law, the first law of thermodynamics, and mechanical equilibrium. Not all topics were included in both tests.) The H&T questions were based on research findings about common student difficulties with this material, including (1) the belief that different materials will have different temperatures under equal external conditions; (2) the belief that the size of an object is a criterion for temperature; (3) difficulties accounting for heat transfers when samples of different temperature or mass are mixed; and (4) a failure to recognize the constancy of the boiling point of water. The four H&T questions on each test roughly corresponded to the four difficulties mentioned.

The first question on each instrument was designed to probe student understanding of thermal equilibrium. The students were asked to consider three common objects that had been in thermal contact with the same surroundings for a long time (and were thereby allowed to come to the same final temperature). On the pretest, the ambient temperature was below normal room temperature; on the posttest, above room temperature. An answer was considered correct if the student stated that all three objects had reached the same temperature. Incorrect answers given by the students were generally based on ideas that the conductive properties of the materials, their atomic structures, or their densities would result in different final temperatures.

The second question concerned the relationship between heat transfer, mass, and temperature change for different samples of water. On the pretest, students were asked to compare the amounts of heat transferred to two different-size samples of water undergoing different changes in temperature. The corresponding question on the posttest involved the mixing of two samples of water. Students were asked to determine the initial temperature of one of the samples from the data given. On the pretest, the students were expected to respond that the amount of heat transferred to each cup would be the same. Correct reasoning indicated an understanding that the amount of heat transferred is proportional to both the mass and the temperature change of each sample. For a correct response to the posttest, students also needed to understand that the amounts of heat transferred to each of the two samples involved in the mixing process are the same in absolute value. (However, since the problem statement included that "no heat can be transferred into or out of the container" we did not expect this additional step to raise the level of difficulty of the question.)

The third question on each test was intended to probe further the students' understanding of heat transfer, focusing on situations in which samples of water of unequal size were mixed. On the pretest, students were expected to predict an approximate final temperature of the mixture. (In particular, students were prompted to indicate whether the final temperature would be greater than, less than, or equal to the arithmetic mean of the initial temperatures.) On the corresponding posttest question, the students were asked to determine the mass of one of the samples (given both initial and final temperatures, and the mass of the other sample). It was expected that students would demonstrate an understanding of the relationship between the mass of the sample and its temperature change. Instead, many students incorrectly reasoned that heat transfer could be determined by simply adding or averaging the given temperatures to arrive at the answer or by describing the heat transfer as being based only on the mass of the sample without consideration of the initial temperature.

The fourth question on 'heat and temperature' concerned two samples of boiling water. One was described as "boiling fast;" the other as "at a slow boil." The posttest question described two samples of boiling water with different masses. In both cases, students were asked to compare the temperature of the two samples. An answer to either question was considered correct, if the student stated that the temperatures of the two samples were equal.

The H&T questions included in the pre- and post-tests are shown in the appendix.

Instruction

Guided-inquiry instruction

The instructional activities carried out by the 'guided-inquiry' group were adapted from the Heat and Temperature module of Physics By Inquiry. Only a subset of the activities in the module was used. The activities chosen were modified to make them suitable for use at the high-school level. In particular, changes were made to match time and equipment constraints, the coverage suggested by the New York State Education Department syllabus, as well as the level of depth that seemed appropriate for the age of the students.

Below, we give a brief description of the activities completed by the 'guided-inquiry' group. A total of ten 40-minute periods was allotted for instruction on heat and temperature. (There was no additional lecture or laboratory instruction on this topic.)

Measurements of temperature

An important goal of the first group of activities was to help students understand temperature as an operationally defined quantity. In addition, students were expected to recognize that the temperature of an object is equal to the temperature of any of its parts, and that the temperature of boiling water is a constant (at normal atmospheric pressure). By emphasizing the concept of temperature as an outcome of a measurement process, we hoped to help students overcome difficulties with this concept that may have arisen from intuitive ideas or a misunderstood definition in terms of microscopic entities.

The students were initially asked to make predictions about the temperature of common objects they were given. Following their predictions, the students were asked to measure the temperature of these objects and record their findings. The students found that the temperature of all objects were the same within the accuracy of their measurements. In a second experiment, the students heated water from room temperature to boiling and recorded their observations. In particular, students were asked to state if the temperature of the water changed after the onset of boiling. Finally, students divided one sample of water into different portions (in separate containers). They compared the temperatures of the individual portions to the initial temperature of the entire sample and found that they were the same.

Changes in temperature

In this sequence of experiments, students investigated the relationship between the masses of water samples that were mixed and their respective changes in temperature. By allowing students to observe the equality of ($m|\Delta T|$) for the two samples being mixed, we sought to help them form a distinction between heat transfer and (change in) temperature. (Although this distinction could be phrased in terms of extensive and intensive quantities such terminology was not introduced.)

The students first predicted the final temperatures for mixtures of equal masses of water with different initial temperatures. Each student checked his or her predictions experimentally. The students were then asked to predict the final temperatures of mixtures of different masses of water. Once again, these predictions were checked against experimental results.

Heat capacity and specific heat

In this section, students studied the effect of the type of material as well as the effect of different masses of materials on temperature change. Students first placed equal-mass samples of different metals at 100^{o}C in equal mass samples of water at room temperature. The change in temperature was recorded for each of the metal samples. This experiment was later repeated with different initial temperatures of the metal samples. The intent of these activities was to develop the concept of heat capacity. Next, students investigated the effect of mass on heat transfer by placing different masses of iron of the same initial temperature into equal masses of water. The temperature change for each mass was recorded. The results of these and the previous experiments were compared to develop an understanding of the difference between heat capacity and specific heat.

Phase changes

In this sequence of activities, the students first predicted the effect of mass on the temperature at which water would boil. The students brought samples of water of different mass to boil and recorded the temperature at which boiling occurred. The students then revisited their predictions to compare them to the experimental results. The experimental results showed that a variation in the mass of the water did not affect the temperature at which water would reach its boiling point. Students were next asked to predict what effect the amount of time a sample was boiling would have on the final temperature of the water. Students conducted the experiment and compared the results to their predictions. The results showed that the amount of time for which a water sample had been boiling had no effect on temperature.

Traditional lecture- and laboratory-based instruction

The control group was taught the same content in a lecture-based format. Over the two-week duration of the Internal Energy unit, the students participated in a total of 10 lecture sessions of 40 minutes each and an additional 160 minutes of laboratory experience addressing this content. The laboratory portion for this group consisted of two units: one on thermal expansion, the other on the specific heat of metals. Due to the separate laboratory instruction, the total time of instruction for the traditionally taught group was about 40% more than that for the 'guided-inquiry' group.

Results

The fraction of students answering correctly each of the four H&T pretest questions ranged from about 20% to about 80% for the 'traditionally-taught' group, and from about 15% to about 65% for the 'guided-inquiry' group. For both groups, the 'temperature' question (the first of the four discussed above) seemed to present the greatest difficulty. The respective success rates on individual questions were similar for the two groups; the largest difference observed was about 15 percentage points. Any small differences observed would favor the traditionally taught group but, with the exception of one item (Mixing), cannot be considered significant at the level of an individual question or for the four-item set. [To check for significance, we performed chi-square tests ($p = 0.05$) for each item and for the set of four H&T items.) The majority of students who answered each question correctly also gave (approximately) correct reasoning to support their answer, with a maximum discrepancy of about 15%.

On the post-test, success rates ranged from about 15% to about 85% for the traditionally taught group and from about 60% to about 95% for the guided-inquiry group. On all four items, the 'guided-inquiry' group performed at a (significantly) higher level than the traditionally taught group. For three of the four items, the difference was 35% or greater; for one item (Temperature), the difference was almost 85%. If correct reasoning is taken as the criterion for a correct answer, the difference between the two groups increases even further for three of the four questions.

Tables 1 through 4 (see next page) illustrate the comparison of students' pretest and posttest responses to the questions posed. The frequencies of incorrect reasoning shown in the tables refer to the categories of incorrect reasoning discussed previously in this article. The percentages of correct, incorrect, and blank responses are printed in bold face and add up to 100%.

Discussion

The data presented here suggest that physics instruction using an inquiry-based approach can be considerably more effective than a traditional approach in enabling students to answer qualitative and quantitative questions about the course content. Furthermore, students taught in a process of guided inquiry also seem to be more likely to overcome common conceptual difficulties and to develop sound reasoning skills. It is therefore, extremely important that we model such approaches in preparing our preservice physics teachers.

It is important to view the results of the reported study in light of the implications that they may have in terms of preparing future physics teachers. In his address to the American Association of Colleges of Teacher Education in February 2000, Arturo Pacheco encouraged teacher education faculty members to embrace change as they consider the most effective ways to prepare future teachers. He recommended that reform efforts, which lead to the simultaneous renewal of public schools and university preparation programs, must be the priority of teacher educators. Most importantly, Pacheco reminded his audience "better teachers lead to better schools. Better schools lead to better children." (p. 8). In a time when accountability and performance are topics at the forefront of the discussions among education leaders, the philosophical stance expressed by Pacheco must become the force behind the preparation of teachers. Until we as teacher educators are willing to share accountability for student learning, the success of teacher education reform and the usefulness of performance-based preparation will not be realized.

A widely accepted model for accomplishing such a goal is that of a reflective practitioner (Schon, 1983; 1987) whose thinking begins with the search for reasonable grounds for belief and action, is assessed on the basis of consequences that result, and is revised accordingly. Skilled professionals monitor, analyze, and adjust their behavior as a function of both its underlying rationale and its consequences. In such instances, reflection effectively integrates theory and practice and places practice in a larger context of meaning while simultaneously focusing theory on the attainment of concrete results (Brell, Caravella, et al., 1999). The format and structure of the reported study demonstrate an example of such a practice being used.

Professional reflection and responsiveness to pupil performance is equally evident in the work of "expert" classroom teachers and applied educational researchers (e.g., Corno & Snow,

Table 1

	Pre-test		Post-test	
Temperature	Traditional (N = 55)	Inquiry-based (N = 74)	Traditional (N = 55)	Inquiry-based (N = 74)
Correct response	**20%**	**15%**	**13%**	**97%**
with correct reasoning	20%	12%	11%	81%
Incorrect responses	**78%**	**74%**	**82%**	**3%**
based on ‘atomic structure’	16%	11%	49%	0%
based on conductive properties	24%	45%	29%	1%
based on density	13%	8%	5%	0%
No response	**2%**	**11%**	**5%**	**0%**

Table 2

	Pre-test		Post-test	
Heat Transfer	Traditional (N = 55)	Inquiry-based (N = 74)	Traditional (N = 55)	Inquiry-based (N = 74)
Correct response	**56%**	**49%**	**33%**	**74%**
with correct reasoning	51%	49%	20%	73%
Incorrect responses	**44%**	**51%**	**67%**	**23%**
Q assumed independent of mass	13%	9%	9%	1%
Q assumed independent of _T	20%	38%	NA	NA
No response	**0%**	**0%**	**0%**	**3%**

Table 3

	Pre-test		Post-test	
Mixing	Traditional (N = 55)	Inquiry-based (N = 74)	Traditional (N = 55)	Inquiry-based (N = 74)
Correct response	**82%**	**65%**	**27%**	**62%**
with correct reasoning	82%	51%	18%	62%
Incorrect responses	**18%**	**35%**	**69%**	**35%**
Average of initial temperatures	15%	11%	NA	NA
Incorrect accounting of mass	3%	24%	45%	15%
No response	**0%**	**0%**	**4%**	**3%**

Table 4

	Pre-test		Post-test	
Boiling	Traditional (N = 55)	Inquiry-based (N = 74)	Traditional (N = 55)	Inquiry-based (N = 74)
Correct response	**49%**	**51%**	**85%**	**96%**
with correct reasoning	38%	35%	76%	95%
Incorrect responses	**47%**	**46%**	**15%**	**3%**
T depends on ‘kinetic energy’	22%	28%	5%	3%
T depends on mass	0%	0%	9%	0%
No response	**4%**	**3%**	**0%**	**1%**

1986; Martella, Nelson, Marchand-Martella, 1999). Here, professionals identify problems or questions in their work, develop plans to address these concerns, implement their plans, and then rigorously evaluate the effects of their actions. In the case of classroom teachers, such reflection is referred to as practical inquiry and is conducted to improve one's work life. These professional reflections often result in the retention of instructional practices that promote pupil progress and the adaptation or discarding of less effective teaching strategies (Corno & Snow, 1986). Applied educational researchers, by the same token, use noticeable improvements in pupil performance as the basis for validating basic scientific principles and the teaching practices they embody. In other professional disciplines (e.g., medicine, engineering, and physics), for example, it is as common for practitioners' findings to contribute theoretically to the scientific community as it is for the findings of researchers to improve practice (Greenwood & Maheady, 1997; Levine, 1995; Vincenti, 1990). Our intent should be to establish a similar relationship within the context of preparing future physics teachers by helping our candidates "bridge the gap" between theory, research, and practice in their daily instructional interactions.

To date, however, two aspects of the reflective practitioner model have remained under-evaluated. These aspects include: (a) the relationship between professional reflection and instructional behavior, and (b) the resultant impact of changes in teaching practice on pupil outcomes (Howey, 1996). In other words, to what extent do professionals' reflections influence what and how they teach? And, perhaps more importantly, what impact do any subsequent changes in instructional practice (as a result of professional reflection) have on pupil performance? We believe that reflection without responsiveness and responsiveness in the absence of improved learner performance is unacceptable at any level of educational reform. As such, questions regarding the relationships among professional reflection, instructional practice, and pupil outcomes should serve as a basis for the teacher preparation program. Programs that connect the effects of teaching to the academic growth of students (Sanders, Saxton, & Horn, 1997) are a logical next step for the assessment of our teacher preparation programs. Given the constant criticism being voiced about teacher preparation and the disapproval that is so frequently expressed about the quality of public school education, teacher educators would be well served by embracing and expanding upon professional practices that openly demonstrate not only the knowledge and skills being mastered by their candidates, but the impact of their programs graduates instruction on student learning.

Focusing on issues such as competency and effectiveness when considering the impact teachers have on student learning, necessitates that teacher educators use performance-based learning in their own classrooms. For example, in the fall of 2000, the National Council for the Accreditation of Teacher Education (NCATE) began piloting performance based accreditation standards for teacher education programs. The foundation of these new accreditation standards is the belief that teacher candidates should possess certain types of knowledge and be able to demonstrate specific skills and behaviors as part of their preparation and training (NCATE, 2000).

Also at the heart of our teacher preparation programs should be the fundamental belief that learning (at all levels) results from the dynamic interplay between pupil and teachers. As candidates teach, they also learn. They learn more about the content they teach, the students they instruct, and their own abilities to teach more effectively. In essence, good teaching practice becomes a recursive process that involves "informed" professional judgment and decision-making.

On the basis of the pretest results of the study reported we conclude that any differences in the initial level of preparation of the two groups are small and would favor the traditionally taught group. Evaluation of student responses to the posttest questions shows sizeable improvement in the frequency of both correct responses and correct reasoning for the 'guided-inquiry' group while improvement for the 'traditionally taught' group, if any, is much smaller. For the 'guided-inquiry' group, there is also a substantial decrease in the frequency of specific incorrect responses that seem to result from the student difficulties described in the literature. If we train our preservice candidates to make classroom decisions informed by students' responsiveness to instruction then effective teacher candidates will retain instructional strategies to which their students are most responsive, while discarding or adapting those that do not benefit learners.

Practices such as those in the study presented can serve as preliminary methods by which to measure candidate performance. It stands to reason, however, given the emphasis on accountability, that a more direct connection between P-12 student performance and the quality of teacher preparation will be emphasized in the future. The inquiry-based classroom approach described is not traditionally what our preservice physics teachers' experience either in their own secondary education or in their preparation for entering the profession. We realize that while we hope that our preservice teachers will adopt such an approach in their own classrooms, these preservice teachers are most likely to teach in a manner in which they have been taught. It therefore stands to reason that we should train our preservice physics teachers using methods and strategies that we would like to see brought into practice.

Acknowledgments

We would like to thank the teacher of the 'traditionally taught' student group for participating in the original study described in this article.

References

Brell, D., Caravella, T., Dell, S., Heist, J., Hennen, F., Sweeney, W., Tibbetts, R., Williams, C., & Wollman-Bonilla, J. (1999, January). *Developing reflective practitioners: A conceptual framework.* Providence, RI: Rhode Island College, Feinstein School of Education and Human Development.

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook for Research on Teaching* (3rd Ed.) (pp. 605-629). New York: Simon & Schuster Macmillan.

Driver R, Squires A, Rushworth P and Wood-Robinson V (1994). *Making Sense of Secondary Science* (London: Routledge)

Erickson G. (1979). Children's conceptions of heat and temperature. *Science Education, 63,* 221-30.

Erickson G. & Tiberghien A. (1985) Heat and Temperature. *Children's Ideas in Science,* ed. R Driver, E Guesne and A. Tiberghien (Milton Keynes: Open University Press).

Greenwood, C. R., & Maheady, L. (1997). Measurable change in student performance: Forgotten standard in teacher preparation? *Teacher Education and Special Education, 20,* 265-275.

Howey, K. (1996). Designing coherent and effective teacher education programs. In J. Sikula, T. J. Buttery, & E. Guyton (Eds.), *Handbook for research on teacher education* (pp. 143-169). New York: Macmillan.

Levine, M. (1995). 21st century professional education: How education could learn from medicine, business, and engineering. *Education Week, 14* (9), 33, 36.

Martella, R.C., Nelson, R., & Marchand-Martella, N. E. (1999). *Research methods: Learning to become a critical research consumer.* Boston: Allyn & Bacon.

McDermott L and the Physics Education Group at the University of Washington (1996). *Physics By Inquiry* (New York: Wiley).

New York State Education Department (1987). *Regents Physics Syllabus* (Albany: The University of the State of New York)

Nachmias R, Stavy R and Avrams R (1990). A microcomputer-based diagnostic system for identifying students' conception of heat and temperature. *Int. J. Sci. Educ., 12,* 123-32.

National Council for Accreditation of Teacher Education (2000). *NCATE 2000 Standards.* Washington, D.C.: Author.

Pacheco, A. (February, 2000). Meeting the challenge of high-quality teacher education: Why higher education must change. Presentation at the annual meeting of the American Association of Colleges of Teacher Education, Chicago, IL.

Rosenquist M (1982). *Improving preparation for college physics of minority students aspiring to science-related careers* Unpublished dissertation (University of Washington).

Sanders, W., Saxton, A., & Horn, S. (1997). The Tennessee value-added assessment system: A challenge to familiar assessment methods. In J. Millman (Ed.), *Grading Teachers, Grading Schools* (pp. 137-162). Thousand Oaks, CA: Corwin Press.

Schon, D. (1983). *The Reflective Practitioner.* New York: Basic Books.

Schon, D. (1987). *Educating the Reflective Practitioner: Toward a Design for Teaching and Learning in the Professions.* San Francisco: Jossey-Bass.

Stavy, R. & Berkovitz, B. (1980). Cognitive conflict as a basis for teaching quantitative aspects of the concept of temperature. *Science Education, 64,* 679-92.

Sokoloff, D. & Thornton, R. (1997). Pre-Test on Heat and Temperature, *Tools for Scientific Thinking,* Medford: CSMT Tufts University.

Vincenti, W. O. (1990). *What Engineers Know and How They Know It.* Baltimore, MD: The Johns Hopkins University Press.

APPENDIX:

PRETEST QUESTIONS	POSTTEST QUESTIONS
Concept: Temperature Three objects have been kept outside for a long time on a cold day: a piece of cotton, a piece of wood and a piece of metal. Which of the objects above has the lowest temperature? If any of the temperatures are the same, state so explicitly. Explain your reasoning.	**Concept:** Temperature An elementary student performs an experiment where she places a book, a wooden block, and a metal ruler in a warm oven that has been turned off. After they stay there over night, which of these objects would be the warmest? The coldest? Explain your reasoning.
Concept: Heat Transfer Cup A contains 100 grams of water. Cup B contains twice as much water. Initially, the water in both cups was at 25°C. Cup A was then heated to 75°C and cup B was heated to 50°C. Which cup had more heat transferred to it? If the amount of heat transferred to both cups is the same, state so explicitly. Explain your reasoning.	**Concept:** Heat Transfer Cup C contains 100 grams of water at 10°C. Cup D, contains 200 grams of water at an unknown temperature. The contents of the two cups are mixed together in an insulated container. (No heat can be transferred into or out of the container.) The final temperature is 50°C. What was the initial temperature of the water in cup D? Explain your reasoning.
Concept: Mixing Cup A now contains 100 grams of **water at 0°C.** Cup B contains 200 grams of water at 50°C. The contents of the two cups are mixed together in an insulated container. (No heat can be transferred into or out of the container.) The final temperature of the water in the container is **A)** lower than 0°C **B)** 0°C **C)** between 0°C and 25°C **D)** 25°C **E)** between 25°C and 50°C **F)** 50°C **G)** higher than 50°C Explain your reasoning.	**Concept:** Mixing Cup E contains 100 grams of water at 40°C. When it was mixed with water in a second cup, cup F at 80°C, the final temperature was 70°C. What was the mass of the water in the second cup before mixing? Explain your reasoning.
Concept: Boiling Point Two pots of boiling water are on a stove. In pot 1, the water is boiling very fast; in pot 2, it is at a "slow boil." Is the temperature of pot 2 *higher than, lower than,* or *the same as* the temperature in pot 1? Explain your reasoning.	**Concept:** Boiling Point A student puts 1000 mL of water in a beaker, and places it on a burner bringing it to a boil. A second student brings 50 mL of water in a beaker to a boil. Is the temperature of the water after it is boiling in the 1000-mL beaker *higher than, lower than,* or *the same as* the temperature of the water in the 50-mL beaker after it is boiling? Explain your reasoning.

